



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0066119
(43) 공개일자 2020년06월09일

(51) 국제특허분류(Int. Cl.)
G06Q 50/10 (2012.01) G06N 3/02 (2019.01)
(52) CPC특허분류
G06Q 50/10 (2013.01)
G06N 3/02 (2019.01)
(21) 출원번호 10-2019-0008303
(22) 출원일자 2019년01월22일
심사청구일자 2019년01월22일
(30) 우선권주장
1020180151971 2018년11월30일 대한민국(KR)

(71) 출원인
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
안예찬
경기도 안양시 동안구 안양판교로 42, 110동 2101호 (관양동, 인덕원마을삼성아파트)
정창성
서울특별시 노원구 노원로 214, 6동 905호 (하계동, 삼익선경아파트)
(74) 대리인
윤귀상

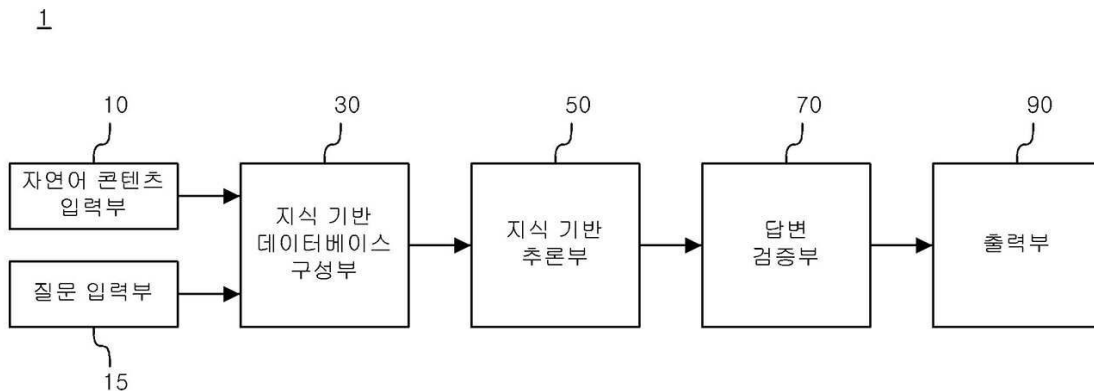
전체 청구항 수 : 총 16 항

(54) 발명의 명칭 지식 기반 추론 및 신뢰도 분석을 이용한 가짜뉴스 판별 방법, 이를 수행하기 위한 기록매체 및 장치

(57) 요약

지식 기반 추론 및 신뢰도 분석을 이용한 가짜뉴스 판별 방법, 이를 수행하기 위한 기록매체 및 장치가 개시된다. 지식 기반 추론 및 신뢰도 분석을 이용한 가짜뉴스 판별 장치는 판별 대상 뉴스에 해당하는 질문을 입력 받는 질문 입력부, 인터넷 상의 자연어로 이루어진 문장에 해당하는 자연어 콘텐츠를 입력 받는 자연어 콘텐츠 입력부, 상기 자연어 콘텐츠를 분석하여 지식 기반 데이터베이스를 구성하고, 상기 지식 기반 데이터베이스로부터 상기 질문과 관련된 자연어 콘텐츠를 상기 질문에 대한 답변 후보군으로 추출하는 지식 기반 데이터베이스 구성부 및 상기 질문 및 상기 답변 후보군을 상기 질문의 형태에 따라 정해지는 추론 모델에 입력하여 상기 질문에 대한 답변을 생성하는 지식 기반 추론부를 포함한다.

대표도 - 도1



이 발명을 지원한 국가연구개발사업

과제고유번호 1711073495

부처명 과학기술정보통신부

연구관리전문기관 정보통신기술진흥센터

연구사업명 인공지능산업원천기술개발

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

판별 대상 뉴스에 해당하는 질문을 입력 받는 질문 입력부;

인터넷 상의 자연어로 이루어진 문장에 해당하는 자연어 콘텐츠를 입력 받는 자연어 콘텐츠 입력부;

상기 자연어 콘텐츠를 분석하여 지식 기반 데이터베이스를 구성하고, 상기 지식 기반 데이터베이스로부터 상기 질문과 관련된 자연어 콘텐츠를 상기 질문에 대한 답변 후보군으로 추출하는 지식 기반 데이터베이스 구성부; 및

상기 질문 및 상기 답변 후보군을 상기 질문의 형태에 따라 정해지는 추론 모델에 입력하여 상기 질문에 대한 답변을 생성하는 지식 기반 추론부;를 포함하는, 지식 기반 추론 및 신뢰도 분석을 이용한 가짜뉴스 판별 장치.

청구항 2

제1항에 있어서,

상기 지식 기반 추론부는,

상기 질문의 형태를 분석하여 상기 질문에 대한 답변 형태를 참/거짓 문답의 답변, 인과 답변 및 확인 답변 중 하나로 설정하는 답변 형태 분석부;를 포함하는, 지식 기반 추론 및 신뢰도 분석을 이용한 가짜뉴스 판별 장치.

청구항 3

제2항에 있어서,

상기 지식 기반 추론부는,

상기 질문에 대한 답변 형태가 참/거짓 문답의 답변으로 설정되는 경우, 상기 질문 및 상기 답변 후보군을 트리화하여 상기 질문과 상기 답변 후보군에 포함되는 자연어 콘텐츠의 의도를 비교하는 트리구조 분석 모델을 상기 질문의 추론 모델로 설정하고,

상기 질문에 대한 답변 형태가 인과 답변 또는 확인 답변으로 설정되는 경우, 딥러닝 학습 기반의 추론 모델을 상기 질문의 추론 모델로 설정하는 추론 엔진 모델 선택부;를 더 포함하는, 지식 기반 추론 및 신뢰도 분석을 이용한 가짜뉴스 판별 장치.

청구항 4

제1항에 있어서,

상기 지식 기반 데이터베이스 구성부는,

상기 자연어 콘텐츠를 도메인의 스코프, 미디어 종류 및 도메인의 카테고리에 따라 분류하고, 상기 자연어 콘텐츠를 데이터의 크기, 언어 형태 및 데이터의 구조에 따라 분류하며, 상기 자연어 콘텐츠의 문장 구조를 분석하여 벡터 값을 산출하는 입력 데이터 분석부;를 포함하는, 지식 기반 추론 및 신뢰도 분석을 이용한 가짜뉴스 판별 장치.

청구항 5

제4항에 있어서,

상기 지식 기반 데이터베이스 구성부는,

상기 자연어 콘텐츠의 분류 결과에 따라 상기 자연어 콘텐츠의 전달 루트를 분석하여 상기 자연어 콘텐츠의 출처 경위를 확인하고, 상기 자연어 콘텐츠를 출처 경위에 따라 계층적으로 분류하여 상기 지식 기반 데이터베이스로 저장하는 계층적 출처 분석부;를 더 포함하는, 지식 기반 추론 및 신뢰도 분석을 이용한 가짜뉴스 판별 장

치.

청구항 6

제5항에 있어서,

상기 지식 기반 데이터베이스 구성부는,

상기 질문의 출처를 식별하고, 상기 질문의 출처에 기반하여 상기 지식 기반 데이터베이스로부터 상기 질문과 관련된 자연어 콘텐츠를 상기 질문에 대한 답변 후보군으로 추출하는 신뢰성 메트릭 기반 분석부;를 더 포함하는, 지식 기반 추론 및 신뢰도 분석을 이용한 가짜뉴스 판별 장치.

청구항 7

제1항에 있어서,

상기 지식 기반 추론부는,

상기 추론 모델 중 하나로 상기 질문 및 상기 답변 후보군의 문장 구조를 트리화하고, 상기 질문 및 상기 답변 후보군에 대한 개체명 및 구문론적 분석을 수행하여 상기 질문과 상기 답변 후보군에 포함되는 자연어 콘텐츠의 의도를 비교하는 방식으로 상기 질문에 대한 답변을 생성하는 트리구조 분석 모델; 및

상기 답변 후보군의 문장 구조를 분석하여 산출되는 벡터 값을 BERT 딥러닝 네트워크에 입력하여 상기 질문에 대한 답변을 생성하는 딥러닝 학습 기반 추론 모델;을 포함하는, 지식 기반 추론 및 신뢰도 분석을 이용한 가짜뉴스 판별 장치.

청구항 8

제1항에 있어서,

상기 지식 기반 추론부에서 생성하는 답변의 신뢰성을 검증하는 답변 검증부;를 더 포함하는, 지식 기반 추론 및 신뢰도 분석을 이용한 가짜뉴스 판별 장치.

청구항 9

판별 대상 뉴스에 해당하는 질문을 입력 받는 단계;

인터넷 상의 자연어로 이루어진 문장에 해당하는 자연어 콘텐츠를 입력 받는 단계;

상기 자연어 콘텐츠를 분석하여 지식 기반 데이터베이스를 구성하는 단계;

상기 지식 기반 데이터베이스로부터 상기 질문과 관련된 자연어 콘텐츠를 상기 질문에 대한 답변 후보군으로 추출하는 단계; 및

상기 질문 및 상기 답변 후보군을 상기 질문의 형태에 따라 정해지는 추론 모델에 입력하여 상기 질문에 대한 답변을 생성하는 단계;를 포함하는, 지식 기반 추론 및 신뢰도 분석을 이용한 가짜뉴스 판별 방법.

청구항 10

제9항에 있어서,

상기 자연어 콘텐츠를 분석하여 지식 기반 데이터베이스를 구성하는 단계는,

상기 질문의 형태를 분석하여 상기 질문에 대한 답변 형태를 참/거짓 문답의 답변, 인과 답변 및 확인 답변 중 하나로 설정하는 단계;를 포함하는, 지식 기반 추론 및 신뢰도 분석을 이용한 가짜뉴스 판별 방법.

청구항 11

제10항에 있어서,

상기 질문 및 상기 답변 후보군을 상기 질문의 형태에 따라 정해지는 추론 모델에 입력하여 상기 질문에 대한 답변을 생성하는 단계는,

상기 질문에 대한 답변 형태가 참/거짓 문답의 답변으로 설정되는 경우, 상기 질문 및 상기 답변 후보군을 트리

화하여 상기 질문과 상기 답변 후보군에 포함되는 자연어 콘텐츠의 의도를 비교하는 트리구조 분석 모델을 상기 질문의 추론 모델로 설정하는 단계; 및

상기 질문에 대한 답변 형태가 인과 답변 또는 확인 답변으로 설정되는 경우, 딥러닝 학습 기반의 추론 모델을 상기 질문의 추론 모델로 설정하는 단계;를 더 포함하는, 지식 기반 추론 및 신뢰도 분석을 이용한 가짜뉴스 판별 방법.

청구항 12

제9항에 있어서,

상기 자연어 콘텐츠를 분석하여 지식 기반 데이터베이스를 구성하는 단계는,

상기 자연어 콘텐츠를 도메인의 스코프, 미디어 종류 및 도메인의 카테고리에 따라 분류하는 단계;

상기 자연어 콘텐츠를 데이터의 크기, 언어 형태 및 데이터의 구조에 따라 분류하는 단계; 및

상기 자연어 콘텐츠의 문장 구조를 분석하여 벡터 값을 산출하는 단계;를 포함하는, 지식 기반 추론 및 신뢰도 분석을 이용한 가짜뉴스 판별 방법.

청구항 13

제12항에 있어서,

상기 자연어 콘텐츠를 분석하여 지식 기반 데이터베이스를 구성하는 단계는,

상기 자연어 콘텐츠의 분류 결과에 따라 상기 자연어 콘텐츠의 전달 루트를 분석하여 상기 자연어 콘텐츠의 출처 경위를 확인하는 단계;

상기 자연어 콘텐츠를 출처 경위에 따라 계층적으로 분류하여 상기 지식 기반 데이터베이스로 저장하는 단계;를 더 포함하는, 지식 기반 추론 및 신뢰도 분석을 이용한 가짜뉴스 판별 방법.

청구항 14

제13항에 있어서,

상기 지식 기반 데이터베이스로부터 상기 질문과 관련된 자연어 콘텐츠를 상기 질문에 대한 답변 후보군으로 추출하는 단계는,

상기 질문의 출처를 식별하고, 상기 질문의 출처에 기반하여 상기 지식 기반 데이터베이스로부터 상기 질문과 관련된 자연어 콘텐츠를 상기 질문에 대한 답변 후보군으로 추출하는 단계;를 포함하는, 지식 기반 추론 및 신뢰도 분석을 이용한 가짜뉴스 판별 방법.

청구항 15

제9항에 있어서,

상기 질문 및 상기 답변 후보군을 상기 질문의 형태에 따라 정해지는 추론 모델에 입력하여 상기 질문에 대한 답변을 생성하는 단계는,

상기 추론 모델 중 하나로 상기 질문 및 상기 답변 후보군의 문장 구조를 트리화하고, 상기 질문 및 상기 답변 후보군에 대한 개체명 및 구문론적 분석을 수행하여 상기 질문과 상기 답변 후보군에 포함되는 자연어 콘텐츠의 의도를 비교하는 방식으로 상기 질문에 대한 답변을 생성하는 단계; 및

상기 답변 후보군의 문장 구조를 분석하여 산출되는 벡터 값을 BERT 딥러닝 네트워크에 입력하여 상기 질문에 대한 답변을 생성하는 단계;를 포함하는, 지식 기반 추론 및 신뢰도 분석을 이용한 가짜뉴스 판별 방법.

청구항 16

제9항에 따른 지식 기반 추론 및 신뢰도 분석을 이용한 가짜뉴스 판별 방법을 수행하기 위한, 컴퓨터 프로그램이 기록된 컴퓨터로 판독 가능한 기록 매체.

발명의 설명

기술분야

[0001] 본 발명은 지식 기반 추론 및 신뢰도 분석을 이용한 가짜뉴스 판별 방법, 이를 수행하기 위한 기록매체 및 장치에 관한 것으로서, 더욱 상세하게는 질문에 해당하는 문장의 사실성을 판단하는 지식 기반 추론 및 신뢰도 분석을 이용한 가짜뉴스 판별 방법, 이를 수행하기 위한 기록매체 및 장치에 관한 것이다.

배경기술

[0002] 고속 이동통신망 및 스마트폰이 널리 보급된 모바일 시대가 도래하면서, 각종 사회관계망서비스(SNS)의 사용이 급속도로 증가하고 있다. 특히, 근래에 들어 블로그(blog), 카카오톡(KakaoTalk), 라인(Line), 페이스북(Facebook), 트위터(Twitter), 인스타그램(Instagram) 및 텀블러(Tumblr)와 같은 SNS의 사용이 급증하면서, 각종 SNS를 통한 정보 또는 뉴스의 전달 역시 폭발적으로 증가하고 있다.

[0003] 그러나, 선거와 같은 정치적 행사가 있을 때마다 각종 SNS를 통해 잘못된 정보 또는 뉴스가 퍼지는 경우 또한 적지 않게 발생하는 것이 현실이다. 또한, 더욱 심각한 것은 특정한 목적을 가지고 의도적으로 각종 SNS를 통해 잘못된 정보 또는 뉴스를 전파하는 경우도 빈번히 발생하고 있다는 점인데, 이와 같은 이유로 SNS를 통한 정보 또는 뉴스의 전달이 향후 중요한 사회적 문제가 될 가능성이 높다고 볼 수 있다.

[0004] 한편 위와 같이 특정한 목적을 가지고 의도적으로 조작된 뉴스인 가짜뉴스는 대부분 사람에 의해 탐색되고, 그 진위 여부가 판별된다.

[0005] 예를 들어, 한국등록특허 제10-1869815호에 따르면 SNS 상에 게재된 뉴스에 대해서 SNS 이용자들이 직접 뉴스의 진위 여부를 파악하여 태그를 입력하고, 태그로부터 의심 지수를 산출하는 방법이 개시된다.

[0006] 이처럼 종래의 가짜뉴스 탐색 방법은 사람이 직접 검증하는 방식으로 많은 시간과 노력을 필요로 한다. 따라서, 사람의 판단을 요구하지 않으며, 가짜뉴스를 실시간으로 선별할 수 있는 새로운 가짜뉴스 탐색 모델이 필요하다.

발명의 내용

해결하려는 과제

[0007] 본 발명은 자연어 콘텐츠를 정제한 후 지식 기반 데이터베이스를 구축하고, 지식 기반 데이터베이스를 통해 질문에 대한 답변을 생성하는 지식 기반 추론 및 신뢰도 분석을 이용한 가짜뉴스 판별 방법, 이를 수행하기 위한 기록매체 및 장치를 제공한다.

[0008] 본 발명의 기술적 과제는 이상에서 언급한 기술적 과제로 제한되지 않으며, 언급되지 않은 또 다른 기술적 과제들은 아래의 기재로부터 당업자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

[0009] 상기 과제를 해결하기 지식 기반 추론 및 신뢰도 분석을 이용한 가짜뉴스 판별 장치는 판별 대상 뉴스에 해당하는 질문을 입력 받는 질문 입력부, 인터넷 상의 자연어로 이루어진 문장에 해당하는 자연어 콘텐츠를 입력 받는 자연어 콘텐츠 입력부, 상기 자연어 콘텐츠를 분석하여 지식 기반 데이터베이스를 구성하고, 상기 지식 기반 데이터베이스로부터 상기 질문과 관련된 자연어 콘텐츠를 상기 질문에 대한 답변 후보군으로 추출하는 지식 기반 데이터베이스 구성부 및 상기 질문 및 상기 답변 후보군을 상기 질문의 형태에 따라 정해지는 추론 모델에 입력하여 상기 질문에 대한 답변을 생성하는 지식 기반 추론부를 포함한다.

[0010] 한편, 상기 지식 기반 추론부는, 상기 질문의 형태를 분석하여 상기 질문에 대한 답변 형태를 참/거짓 문답의 답변, 인과 답변 및 확인 답변 중 하나로 설정하는 답변 형태 분석부를 포함할 수 있다.

[0011] 또한, 상기 지식 기반 추론부는, 상기 질문에 대한 답변 형태가 참/거짓 문답의 답변으로 설정되는 경우, 상기 질문 및 상기 답변 후보군을 트리화하여 상기 질문과 상기 답변 후보군에 포함되는 자연어 콘텐츠의 의도를 비교하는 트리구조 분석 모델을 상기 질문의 추론 모델로 설정하고, 상기 질문에 대한 답변 형태가 인과 답변 또는 확인 답변으로 설정되는 경우, 딥러닝 학습 기반의 추론 모델을 상기 질문의 추론 모델로 설정하는 추론 엔진 모델 선택부를 더 포함할 수 있다.

[0012] 또한, 상기 지식 기반 데이터베이스 구성부는, 상기 자연어 콘텐츠를 도메인의 스코프, 미디어 종류 및 도메인

의 카테고리에 따라 분류하고, 상기 자연어 콘텐츠를 데이터의 크기, 언어 형태 및 데이터의 구조에 따라 분류하며, 상기 자연어 콘텐츠의 문장 구조를 분석하여 벡터 값을 산출하는 입력 데이터 분석부를 포함할 수 있다.

[0013] 또한, 상기 지식 기반 데이터베이스 구성부는, 상기 자연어 콘텐츠의 분류 결과에 따라 상기 자연어 콘텐츠의 전달 루트를 분석하여 상기 자연어 콘텐츠의 출처 경위를 확인하고, 상기 자연어 콘텐츠를 출처 경위에 따라 계층적으로 분류하여 상기 지식 기반 데이터베이스로 저장하는 계층적 출처 분석부를 더 포함할 수 있다.

[0014] 또한, 상기 지식 기반 데이터베이스 구성부는, 상기 질문의 출처를 식별하고, 상기 질문의 출처에 기반하여 상기 지식 기반 데이터베이스로부터 상기 질문과 관련된 자연어 콘텐츠를 상기 질문에 대한 답변 후보군으로 추출하는 신뢰성 메트릭 기반 분석부를 더 포함할 수 있다.

[0015] 또한, 상기 지식 기반 추론부는, 상기 추론 모델 중 하나로 상기 질문 및 상기 답변 후보군의 문장 구조를 트리화하고, 상기 질문 및 상기 답변 후보군에 대한 개체명 및 구문론적 분석을 수행하여 상기 질문과 상기 답변 후보군에 포함되는 자연어 콘텐츠의 의도를 비교하는 방식으로 상기 질문에 대한 답변을 생성하는 트리구조 분석 모델 및 상기 답변 후보군의 문장 구조를 분석하여 산출되는 벡터 값을 BERT 딥러닝 네트워크에 입력하여 상기 질문에 대한 답변을 생성하는 딥러닝 학습 기반 추론 모델을 포함할 수 있다.

[0016] 또한, 상기 지식 기반 추론부에서 생성하는 답변의 신뢰성을 검증하는 답변 검증부를 더 포함할 수 있다.

[0017] 한편, 본 발명의 지식 기반 추론 및 신뢰도 분석을 이용한 가짜뉴스 판별 방법은 판별 대상 뉴스에 해당하는 질문을 입력 받는 단계, 인터넷 상의 자연어로 이루어진 문장에 해당하는 자연어 콘텐츠를 입력 받는 단계, 상기 자연어 콘텐츠를 분석하여 지식 기반 데이터베이스를 구성하는 단계, 상기 지식 기반 데이터베이스로부터 상기 질문과 관련된 자연어 콘텐츠를 상기 질문에 대한 답변 후보군으로 추출하는 단계 및 상기 질문 및 상기 답변 후보군을 상기 질문의 형태에 따라 정해지는 추론 모델에 입력하여 상기 질문에 대한 답변을 생성하는 단계를 포함한다.

[0018] 한편, 상기 자연어 콘텐츠를 분석하여 지식 기반 데이터베이스를 구성하는 단계는, 상기 질문의 형태를 분석하여 상기 질문에 대한 답변 형태를 참/거짓 문답의 답변, 인과 답변 및 확인 답변 중 하나로 설정하는 단계를 포함할 수 있다.

[0019] 또한, 상기 질문 및 상기 답변 후보군을 상기 질문의 형태에 따라 정해지는 추론 모델에 입력하여 상기 질문에 대한 답변을 생성하는 단계는, 상기 질문에 대한 답변 형태가 참/거짓 문답의 답변으로 설정되는 경우, 상기 질문 및 상기 답변 후보군을 트리화하여 상기 질문과 상기 답변 후보군에 포함되는 자연어 콘텐츠의 의도를 비교하는 트리구조 분석 모델을 상기 질문의 추론 모델로 설정하는 단계 및 상기 질문에 대한 답변 형태가 인과 답변 또는 확인 답변으로 설정되는 경우, 딥러닝 학습 기반의 추론 모델을 상기 질문의 추론 모델로 설정하는 단계를 더 포함할 수 있다.

[0020] 또한, 상기 자연어 콘텐츠를 분석하여 지식 기반 데이터베이스를 구성하는 단계는, 상기 자연어 콘텐츠를 도메인의 스코프, 미디어 종류 및 도메인의 카테고리에 따라 분류하는 단계, 상기 자연어 콘텐츠를 데이터의 크기, 언어 형태 및 데이터의 구조에 따라 분류하는 단계 및 상기 자연어 콘텐츠의 문장 구조를 분석하여 벡터 값을 산출하는 단계를 포함할 수 있다.

[0021] 또한, 상기 자연어 콘텐츠를 분석하여 지식 기반 데이터베이스를 구성하는 단계는, 상기 자연어 콘텐츠의 분류 결과에 따라 상기 자연어 콘텐츠의 전달 루트를 분석하여 상기 자연어 콘텐츠의 출처 경위를 확인하는 단계, 상기 자연어 콘텐츠를 출처 경위에 따라 계층적으로 분류하여 상기 지식 기반 데이터베이스로 저장하는 단계를 더 포함할 수 있다.

[0022] 또한, 상기 지식 기반 데이터베이스로부터 상기 질문과 관련된 자연어 콘텐츠를 상기 질문에 대한 답변 후보군으로 추출하는 단계는, 상기 질문의 출처를 식별하고, 상기 질문의 출처에 기반하여 상기 지식 기반 데이터베이스로부터 상기 질문과 관련된 자연어 콘텐츠를 상기 질문에 대한 답변 후보군으로 추출하는 단계를 포함할 수 있다.

[0023] 또한, 상기 질문 및 상기 답변 후보군을 상기 질문의 형태에 따라 정해지는 추론 모델에 입력하여 상기 질문에 대한 답변을 생성하는 단계는, 상기 추론 모델 중 하나로 상기 질문 및 상기 답변 후보군의 문장 구조를 트리화하고, 상기 질문 및 상기 답변 후보군에 대한 개체명 및 구문론적 분석을 수행하여 상기 질문과 상기 답변 후보군에 포함되는 자연어 콘텐츠의 의도를 비교하는 방식으로 상기 질문에 대한 답변을 생성하는 단계 및 상기 답변 후보군의 문장 구조를 분석하여 산출되는 벡터 값을 BERT 딥러닝 네트워크에 입력하여 상기 질문에 대한 답

변을 생성하는 단계를 포함할 수 있다.

[0024] 또한, 상기 지식 기반 추론 및 신뢰도 분석을 이용한 가짜뉴스 판별 방법을 수행하기 위한, 컴퓨터 프로그램이 기록된 컴퓨터로 판독 가능한 기록 매체일 수 있다.

발명의 효과

[0025] 본 발명에 따르면 사람이 직접 판별 대상 뉴스를 검증하는 것이 아니라 지식 기반 데이터베이스를 통해 판별 대상 뉴스의 사실성에 대한 답변을 생성하므로 빠르고 정확한 가짜 뉴스 탐색이 가능하다.

도면의 간단한 설명

[0026] 도 1은 본 발명의 일 실시예에 따른 지식 기반 추론 및 신뢰도 분석을 이용한 가짜뉴스 판별 장치의 블록도이다.

도 2는 도 1에 도시된 지식 기반 데이터베이스 구성부의 세부 블록도이다.

도 3은 도 1에 도시된 지식 기반 추론부의 세부 블록도이다.

도 4는 본 발명의 일 실시예에 따른 지식 기반 추론 및 신뢰도 분석을 이용한 가짜뉴스 판별 방법의 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0027] 후술하는 본 발명에 대한 상세한 설명은, 본 발명이 실시될 수 있는 특정 실시예를 예시로서 도시하는 첨부 도면을 참조한다. 이들 실시예는 당업자가 본 발명을 실시할 수 있기에 충분하도록 상세히 설명된다. 본 발명의 다양한 실시예는 서로 다르지만 상호 배타적일 필요는 없음이 이해되어야 한다. 예를 들어, 여기에 기재되어 있는 특정 형상, 구조 및 특성은 일 실시예와 관련하여 본 발명의 정신 및 범위를 벗어나지 않으면서 다른 실시예로 구현될 수 있다. 또한, 각각의 개시된 실시예 내의 개별 구성요소의 위치 또는 배치는 본 발명의 정신 및 범위를 벗어나지 않으면서 변경될 수 있음이 이해되어야 한다. 따라서, 후술하는 상세한 설명은 한정적인 의미로서 취하려는 것이 아니며, 본 발명의 범위는, 적절하게 설명된다면, 그 청구항들이 주장하는 것과 균등한 모든 범위와 더불어 첨부된 청구항에 의해서만 한정된다. 도면에서 유사한 참조부호는 여러 측면에 걸쳐서 동일하거나 유사한 기능을 지칭한다.

[0028] 이하, 도면들을 참조하여 본 발명의 바람직한 실시예들을 보다 상세하게 설명하기로 한다.

[0030] 도 1은 본 발명의 일 실시예에 따른 지식 기반 추론 및 신뢰도 분석을 이용한 가짜뉴스 판별 장치의 블록도이다.

[0031] 도 1을 참조하면, 본 발명의 일 실시예에 따른 지식 기반 추론 및 신뢰도 분석을 이용한 가짜뉴스 판별 장치(1)는 자연어 콘텐츠 입력부(10), 질문 입력부(15), 지식 기반 데이터베이스 구성부(30), 지식 기반 추론부(50), 답변 검증부(70) 및 출력부(90)를 포함한다.

[0032] 본 발명의 일 실시예에 따른 지식 기반 추론 및 신뢰도 분석을 이용한 가짜뉴스 판별 장치(1, 이하 장치)는 자연어 콘텐츠를 이용하여 지식 기반 데이터베이스를 구축하고, 지식 기반 데이터베이스를 토대로 질문 문장에 대해 답변을 생성하여 출력할 수 있다.

[0033] 본 발명의 일 실시예에 따른 장치(1)는 지식 기반 추론 및 신뢰도 분석을 이용하여 질문에 대한 답변 생성을 위한 소프트웨어(어플리케이션)가 설치되어 실행될 수 있으며, 자연어 콘텐츠 입력부(10), 질문 입력부(15), 지식 기반 데이터베이스 구성부(30), 지식 기반 추론부(50), 답변 검증부(70) 및 출력부(90)는 지식 기반 추론 및 신뢰도 분석을 이용하여 질문에 대한 답변 생성을 위한 소프트웨어에 의해 제어될 수 있다.

[0034] 자연어 콘텐츠 입력부(10), 질문 입력부(15), 지식 기반 데이터베이스 구성부(30), 지식 기반 추론부(50), 답변 검증부(70) 및 출력부(90)의 구성은 통합 모듈로 형성되거나, 하나 이상의 모듈로 이루어질 수 있다. 그러나, 이와 반대로 각 구성은 별도의 모듈로 이루어질 수도 있다.

[0035] 본 발명의 일 실시예에 따른 장치(1)는 이동성을 갖거나 고정될 수 있다. 본 발명의 일 실시예에 따른 장치(1)는 컴퓨터(computer), 서버(server) 또는 엔진(engine) 형태일 수 있으며, 디바이스(device), 기구

(apparatus), 단말(terminal), UE(user equipment), MS(mobile station), MT(mobile terminal), UT(user terminal), SS(subscriber station), 무선기기(wireless device), PDA(personal digital assistant), 무선 모뎀(wireless modem), 휴대기기(handheld device) 등 다른 용어로 불릴 수 있다.

- [0036] 이하, 도 1에 도시된 본 발명의 일 실시예에 따른 장치(1)의 각 구성에 대해 구체적으로 설명한다.
- [0037] 자연어 콘텐츠 입력부(10)는 인터넷 상의 자연어로 이루어진 문장에 해당하는 자연어 콘텐츠를 수집할 수 있다.
- [0038] 예를 들면, 자연어 콘텐츠는 인터넷 상의 뉴스, 백과사전, 보고서, 나무위키 등으로부터 수집되는 문장에 해당한다.
- [0039] 질문 입력부(15)는 판별 대상 뉴스에 해당하는 질문 문장을 입력 받을 수 있다.
- [0040] 질문 입력부(15)에서 입력 받는 질문 문장의 형태는 참/거짓 문답의 질문, 인과 질문, 확인 질문, 의견을 구하는 질문 등으로 나뉠 수 있다.
- [0041] 지식 기반 데이터베이스 구성부(30)는 자연어 콘텐츠를 분석하여 지식 기반 데이터베이스를 구성하고, 지식 기반 데이터베이스로부터 질문과 관련된 자연어 콘텐츠를 질문에 대한 답변 후보군으로 추출할 수 있다. 이와 관련하여 구체적인 설명은 도 2를 참조하여 후술한다.
- [0042] 지식 기반 추론부(50)는 질문의 형태에 따라 추론 모델을 선정하고, 질문 및 답변 후보군을 추론 모델에 입력하여 질문에 대한 답변을 생성할 수 있다. 본 실시예에서 추론 모델은 트리구조 분석 모델과 딥러닝 학습 기반의 추론 모델로 나뉠 수 있다. 이와 관련하여 구체적인 설명은 도 3을 참조하여 후술한다.
- [0043] 답변 검증부(70)는 지식 기반 추론부(50)에서 생성하는 답변의 신뢰성을 검증할 수 있다.
- [0044] 예를 들면, 답변 검증부(70)는 이전에 사실로 판별된 문장, 사실로 알려진 문장 등에 답변이 있는지 역 추적하여 답변을 검증할 수 있다.
- [0045] 또는, 답변 검증부(70)는 문장을 출처 경위에 따라 분류하여 저장하는 지식 기반 데이터베이스를 이용하여 답변과 다른 출처에 답변과 비슷한 문장이 있는지를 확인하는 방식으로 답변을 검증할 수 있다.
- [0046] 또는, 답변 검증부(70)는 신뢰성 메트릭을 이용하여 답변을 검증할 수 있다. 신뢰성 메트릭은 지식 기반 데이터베이스 구성부(30)에서 답변 후보군을 추출하기 위해 구성하는 것으로, 시간 분류, 출처 식별, 경로 분석, 대명사 출처, 지식 수준 분석 등의 식별 및 분류 작업이 이루어질 수 있다. 답변 검증부(70)는 신뢰성 메트릭의 시간 분류, 출처 식별, 경로 분석, 대명사 출처, 지식 수준 분석 각각에 해당하는 식별 및 분류에 가중치를 더하여, 답변에 해당하는 신뢰도를 수치적으로 계산할 수도 있다.
- [0047] 출력부(90)는 질문에 대한 답변 문장을 출력할 수 있다.
- [0048] 출력부(90)에서 출력 하는 답변 문장의 형태는 질문의 형태에 따라 참/거짓 문답의 답변, 인과 답변, 확인 답변 등으로 나뉠 수 있다.
- [0050] 도 2는 도 1에 도시된 지식 기반 데이터베이스 구성부의 세부 블록도이다.
- [0051] 도 2를 참조하면, 지식 기반 데이터베이스 구성부(30)는 입력 데이터 분석부(31), 계층적 출처 분석부(33) 및 신뢰성 메트릭 기반 분석부(35)를 포함할 수 있다.
- [0052] 입력 데이터 분석부(31)는 도 1에 도시된 자연어 콘텐츠 입력부(10)에서 수집하는 자연어 콘텐츠를 분석할 수 있다.
- [0053] 입력 데이터 분석부(31)는 자연어 콘텐츠의 도메인, 데이터 타입 및 문장 구조를 분석할 수 있다.
- [0054] 입력 데이터 분석부(31)는 자연어 콘텐츠의 도메인을 분석하여, 자연어 콘텐츠를 도메인 스코프, 미디어의 종류 및 도메인 카테고리에 따라 분류할 수 있다.
- [0055] 입력 데이터 분석부(31)는 자연어 콘텐츠의 도메인 스코프를 정의하고, 자연어 콘텐츠의 미디어 종류를 분류할 수 있다. 예를 들면, 미디어의 종류는 라디오, 텔레비전, 인터넷 뉴스 등에 해당한다. 그리고, 입력 데이터 분석부(31)는 자연어 콘텐츠의 도메인 카테고리를 분류할 수 있다. 예를 들면, 도메인 카테고리는 스포츠, 음악, 정치, IT, 건강 등으로 나뉠 수 있다.

- [0056] 입력 데이터 분석부(31)는 자연어 콘텐츠의 데이터 타입을 분석하여, 자연어 콘텐츠를 데이터의 크기, 언어 형태 및 데이터의 구조에 따라 분류할 수 있다.
- [0057] 입력 데이터 분석부(31)는 자연어 콘텐츠의 데이터 크기를 크게 세 가지 타입으로 분류할 수 있다. 또한 입력 데이터 분석부(31)는 자연어 콘텐츠의 언어 형태를 분류할 수 있다. 예를 들면, 언어 형태의 종류는 한국어, 영어, 스페인어 등에 해당한다. 또한 입력 데이터 분석부(31)는 자연어 콘텐츠의 데이터의 구조를 분류할 수 있다. 예를 들면, 데이터의 구조는 정형 데이터, 준정형 데이터 및 비정형 데이터로 나뉠 수 있다.
- [0058] 입력 데이터 분석부(31)는 자연어 콘텐츠의 문장 구조를 형태론적, 구문론적, 의미론적으로 분석하여 벡터 값을 산출할 수 있다.
- [0059] 입력 데이터 분석부(31)는 자연어 콘텐츠를 최소 의미 단위인 형태소로 나누어 형태론적으로 분석할 수 있다. 입력 데이터 분석부(31)는 자연어 콘텐츠를 문장 단위로 자르고, 각각의 문장마다의 Sen2vec을 이용한 문장 임베딩을 통해 벡터 값으로 나타낼 수 있다. 입력 데이터 분석부(31)는 문장 단위로 자른 자연어 콘텐츠를 Mecab-ko 형태소 분석기를 이용하여 형태소 단위로 나누고, 문맥에서의 단어들 간의 관계를 파악하는 임베딩을 통해 벡터 값으로 나타낼 수 있다. 그리고 입력 데이터 분석부(31)는 형태소의 특징 벡터를 추가할 수 있다. 입력 데이터 분석부(31)는 형태소의 특징으로 형태소의 종류인 형태소의 코드(예를 들면, 인명, 지명, 고유명사, 일반명사, 동사, 기타 등)를 분류하여 각각 가중치를 주고, 형태소의 중요도, 전체 빈도, 문서 내의 빈도, 타이틀 출현, 문서의 카테고리 등을 포함하는 특징 벡터를 추가할 수 있다. 이러한 형태론적 분석을 통해 산출되는 자연어 콘텐츠의 벡터 값은 추후 질문에 대한 답변 후보군 추출에 사용될 수 있을 것이다.
- [0060] 입력 데이터 분석부(31)는 자연어 콘텐츠를 문장을 이루는 구성 성분으로 분해하여 구문론적으로 분석할 수 있다. 입력 데이터 분석부(31)는 자연어 콘텐츠를 그 문장을 이루는 구성 성분으로 분해하고, 구성 성분 간의 위치 관계에 따라 문장의 구조를 결정하여 분류할 수 있다.
- [0061] 입력 데이터 분석부(31)는 자연어 콘텐츠를 작성한 사용자의 의도를 파악하여 의미론적으로 분석할 수 있다. 입력 데이터 분석부(31)는 자연어 콘텐츠의 어휘의 의미와 문맥을 분석하여 자연어 콘텐츠를 해석할 수 있다.
- [0062] 계층적 출처 분석부(33)는 입력 데이터 분석부(31)에서의 자연어 콘텐츠 분석 결과에 따라 자연어 콘텐츠를 서로 연관성 있는 콘텐츠 별로 분류하여 지식 기반 데이터베이스로 저장할 수 있다.
- [0063] 계층적 출처 분석부(33)는 입력 데이터 분석부(31)에서의 자연어 콘텐츠 분류 결과에 따라 자연어 콘텐츠의 전달 루트를 분석하여 해당 내용의 출처 경위를 확인할 수 있다. 계층적 출처 분석부(33)는 자연어 콘텐츠를 출처 경위에 따라 계층적으로 분류하여 지식 기반 데이터베이스로 저장할 수 있다.
- [0064] 계층적 출처 분석부(33)는 자연어 콘텐츠가 서로 다른 출처 경위가 확인되는 경우, 자연어 콘텐츠를 각 출처 경위에 따라 계층적으로 분류하여 저장하되 가중치를 부여할 수 있다.
- [0065] 신뢰성 메트릭 기반 분석부(35)는 지식 기반 데이터베이스에서 질문 입력부(15)에서 입력 받는 질문과 관련된 자연어 콘텐츠를 질문에 대한 답변 후보군으로 추출할 수 있다.
- [0066] 신뢰성 메트릭 기반 분석부(35)는 지식 기반 데이터베이스로부터 답변 후보군을 추출하기 위해 질문을 식별 및 분류하여 신뢰성 메트릭을 구성할 수 있다.
- [0067] 예를 들면, 신뢰성 메트릭 기반 분석부(35)는 시간 분류자를 포함하여 질문이 발생한 시간을 식별할 수 있다. 신뢰성 메트릭 기반 분석부(35)는 질문이 발생한 시간을 이용하여 지식 기반 데이터베이스에서 답변 후보군을 추출할 수 있다.
- [0068] 신뢰성 메트릭 기반 분석부(35)는 출처 식별자를 포함하여 질문의 출처를 식별할 수 있다. 예를 들면, 출처 식별자는 질문에 포함되는 "~에 따르면, ~에 의해"등의 문장으로부터 질문의 출처를 식별할 수 있다. 신뢰성 메트릭 기반 분석부(35)는 질문의 출처를 이용하여 지식 기반 데이터베이스에서 답변 후보군을 추출할 수 있다.
- [0069] 신뢰성 메트릭 기반 분석부(35)는 경로 분석기를 포함하여 지식 기반 데이터베이스에서 질문의 출처 경위를 찾아 실제 질문에 해당하는 자연어 콘텐츠가 존재하는지를 파악할 수 있다. 이러한 신뢰성 메트릭 기반 분석부(35)의 경로 탐색 결과는 질문이 가짜 뉴스에 해당하는지를 판단하거나, 답변의 신뢰성을 검증하는 데에 사용될 수 있다.
- [0070] 신뢰성 메트릭 기반 분석부(35)는 대명사 출처 분석기를 포함하여 질문의 주어에 해당하는 인물, 기관 등을 식별할 수 있다. 대명사 출처 분석기는 주어에 따라 가중치를 부여할 수 있다. 신뢰성 메트릭 기반 분석부(35)는

질문의 주어를 이용하여 지식 기반 데이터베이스에서 답변 후보군을 추출할 수 있다.

- [0071] 신뢰성 메트릭 기반 분석부(35)는 지식 수준 분석기를 포함하여 대명사 출처 분석기에서 식별한 질문의 주어의 지식 수준을 분석할 수 있다. 신뢰성 메트릭 기반 분석부(35)는 지식 기반 데이터베이스를 이용하여 질문의 주어가 관련 지식에 대해 충분한 판단과 결과를 얻어낼 수 있는 사람인지, 해당 부분의 지식과 관련 없는 사람인지를 비교할 수 있다.
- [0073] 도 3은 도 1에 도시된 지식 기반 추론부의 세부 블록도이다.
- [0074] 도 3을 참조하면, 지식 기반 추론부(50)는 답변 형태 분석부(51), 추론 엔진 모델 선택부(53), 데이터 전처리부(55), 트리구조 분석 모델(57) 및 딥러닝 학습 기반 추론 모델(59)을 포함할 수 있다.
- [0075] 답변 형태 분석부(51)는 질문 입력부(15)에서 입력 받는 질문 문장의 형태에 따라 답변 형태를 설정할 수 있다.
- [0076] 답변 형태 분석부(51)는 질문 문장의 형태를 분석하여 질문에 대한 답변 형태를 참/거짓 문답의 답변, 인과 답변 및 확인 답변 중 하나로 설정할 수 있다.
- [0077] 추론 엔진 모델 선택부(53)는 답변 형태에 따라 추론 모델을 설정할 수 있다. 여기서, 추론 모델은 트리구조 분석 모델(57) 또는 딥러닝 학습 기반 추론 모델(59) 중 하나일 수 있다.
- [0078] 추론 엔진 모델 선택부(53)는 답변 형태가 참/거짓 문답의 답변으로 설정되는 경우, 추론 모델로 트리구조 분석 모델(57)을 설정할 수 있다.
- [0079] 추론 엔진 모델 선택부(53)는 답변 형태가 인과 답변 또는 확인 답변과 같이 참/거짓 문답의 답변이 아닌 답변으로 설정되는 경우, 추론 모델로 딥러닝 학습 기반의 추론 모델을 설정할 수 있다.
- [0080] 데이터 전처리부(55)는 신뢰성 메트릭 기반 분석부(35)에서 추출하는 질문에 대한 답변 후보군을 추론 모델의 입력 데이터로 사용할 수 있도록 답변 후보군의 형태론적 분석을 진행할 수 있다.
- [0081] 데이터 전처리부(55)는 답변 후보군에 포함되는 자연어 콘텐츠에 있어서 상술한 입력 데이터 분석부(31)와 같이 형태론적 분석을 재차 진행하여 벡터 값을 산출할 수 있다.
- [0082] 트리구조 분석 모델(57)은 추론 모델 중 하나로 답변 후보군으로부터 답변을 생성할 수 있다.
- [0083] 트리구조 분석 모델(57)은 입력 데이터로 질문 문장 및 답변 후보군의 문장을 사용할 수 있다. 트리구조 분석 모델(57)은 질문 문장 및 답변 후보군의 문장 구조를 트리화하고, 질문 및 답변 후보군에 대한 개체명 및 구문론적 분석을 수행하여 질문과 답변 후보군에 포함되는 자연어 콘텐츠의 의도를 비교하는 방식으로 최종 결과를 추론할 수 있다.
- [0084] 트리구조 분석 모델(57)은 질문 문장 및 답변 후보군의 트리구조에 대한 개체명 분석 방법으로, 문장의 의도를 파악하고 어휘의 의미와 문맥을 분석하여 문장을 해석할 수 있다. 예를 들면, 트리구조 분석 모델(57)은 한국어의 개체명 인식 딥러닝 방식 중 하나인 BiLSTM-CRF를 활용하여 개체명 인식으로 단어 간의 의미를 태깅할 수 있다. 개체명 태깅은 인물(PERSON), 학문분야(STUDY_FIELD), 이론(THEORY), 인공물(ARTIFACTS), 기관(ORGANIZATION), 지역(LOCATION), 문명(CIVILIZATION), 날짜(DATE), 시간(TIME), 수량(QUANTITY), 이벤트(EVENT), 동물(ANIMAL), 식물(PLANT), 물질(MATERIAL), 용어(TERM)의 15개로 나눌 수 있다. 트리구조 분석 모델(57)은 질문에 태깅된 개체명이 답변 후보군에 포함되는 자연어 콘텐츠에 각각 의미론적으로 알맞게 속해있는지를 확인하는 개체명 분석을 수행할 수 있다.
- [0085] 트리구조 분석 모델(57)은 질문 문장 및 답변 후보군의 트리구조에 대한 구문론적 분석 방법으로, 질문 및 답변 후보군의 문장 구성 성분 간의 위치 관계를 분석하여 문장의 구조를 해석할 수 있다. 예를 들면, 트리구조 분석 모델(57)은 문장에서의 각 형태소간의 문장 문법을 파악하여 구문 태그와 기능 태그를 통해 문장의 형태소간의 관계를 파악할 수 있다. 구문 태그로는 NP_체언(명사, 대명사, 수사), VP_용언(동사, 형용사, 보조용언), AP_부사구, VNP_긍정 지정사구(명사+이다), DP_관형사구, IP_감탄사구(호칭 및 대답 등의 표현), X_의사 구(pseudo phrase, 조사 단독 어절 또는 기호 등), L_부호(왼쪽 괄호 및 따옴표), R_부호(오른쪽 괄호 및 따옴표) 등으로 나눌 수 있다. 기능 태그로는 SBJ_주어, OBJ_목적어, MOD_관형어(체언 수식어), AJT_부사어(용언 수식어), CMP_보어, CNJ_접속어(~와) 등으로 나눌 수 있다.
- [0086] 트리구조 분석 모델(57)은 이와 같이 질문 및 답변 후보군의 트리구조에 대한 개체명 및 구문론적 비교 분석을

수행하여 질문의 의도에 해당하는 내용이 답변 후보군에 존재하는지 여부에 따라 참/거짓의 답변을 생성할 수 있을 것이다.

[0087] 딥러닝 학습 기반 추론 모델(59)은 답변 후보군의 문장 구조를 분석하여 산출되는 벡터 값을 BERT 딥러닝 네트워크에 입력하여 질문에 대한 답변을 생성할 수 있다. 예를 들면, 답변 후보군의 문장 구조를 분석하여 단어 임베딩을 통해 산출되는 벡터 값은 300차원이고, 형태소의 특징 벡터 값은 300차원으로 총 600차원의 벡터일 수 있다. BERT 딥러닝 네트워크는 질문 답변 방법의 오픈소스로 공개되어 있으며, 딥러닝 학습 기반 추론 모델(59)은 BERT 딥러닝 네트워크에 600차원의 벡터를 입력 값으로 하여 결과 값을 획득할 수 있으며, 그 결과 값을 질문에 대한 답변으로 생성할 수 있을 것이다.

[0089] 이와 같은, 본 발명의 일 실시예에 따른 장치(1)는 자연어 콘텐츠를 정제한 후 지식 기반 데이터베이스를 구축하고, 지식 기반 데이터베이스를 통해 질문에 대한 답변을 생성할 수 있다. 즉, 본 발명의 일 실시예에 따른 장치(1)는 사람이 직접 판별 대상 뉴스를 검증하는 것이 아니라 지식 기반 데이터베이스를 통해 판별 대상 뉴스의 사실성에 대한 답변을 생성하므로 빠르고 정확한 가짜 뉴스 탐색이 가능하다.

[0091] 이하 도 4를 참조하여 본 발명의 일 실시예에 따른 지식 기반 추론 및 신뢰도 분석을 이용한 가짜뉴스 판별 방법에 대해 설명한다.

[0092] 도 4는 본 발명의 일 실시예에 따른 지식 기반 추론 및 신뢰도 분석을 이용한 가짜뉴스 판별 방법의 흐름도이다.

[0093] 본 발명의 일 실시예에 따른 지식 기반 추론 및 신뢰도 분석을 이용한 가짜뉴스 판별 방법은 도 1에 도시된 본 발명의 일 실시예에 따른 장치(1)와 실질적으로 동일한 구성에서 진행될 수 있다. 따라서 도 1의 장치(1)와 동일한 구성요소는 동일한 도면부호를 부여하고, 반복되는 설명은 생략한다.

[0094] 도 4를 참조하면, 자연어 콘텐츠 입력부(10) 및 질문 입력부(15)는 각각 자연어 콘텐츠 및 질문을 입력 받을 수 있다(S100).

[0095] 자연어 콘텐츠 입력부(10)는 인터넷 상의 자연어로 이루어진 문장에 해당하는 자연어 콘텐츠를 수집할 수 있다. 예를 들면, 자연어 콘텐츠는 인터넷 상의 뉴스, 백과사전, 보고서, 나무위키 등으로부터 수집되는 문장에 해당한다.

[0096] 질문 입력부(15)는 판별 대상 뉴스에 해당하는 질문 문장을 입력 받을 수 있다. 질문 입력부(15)에서 입력 받는 질문 문장의 형태는 참/거짓 문답의 질문, 인과 질문, 확인 질문, 의견을 구하는 질문 등으로 나뉠 수 있다.

[0097] 지식 기반 데이터베이스 구성부(30)는 자연어 콘텐츠를 분석하여 지식 기반 데이터베이스를 구성할 수 있다(S200).

[0098] 지식 기반 데이터베이스 구성부(30)는 자연어 콘텐츠의 도메인, 데이터 타입 및 문장 구조를 분석할 수 있다. 지식 기반 데이터베이스 구성부(30)는 자연어 콘텐츠 분석 결과에 따라 자연어 콘텐츠를 서로 연관성 있는 콘텐츠 별로 분류하여 지식 기반 데이터베이스로 저장할 수 있다. 지식 기반 데이터베이스 구성부(30)는 자연어 콘텐츠 분류 결과에 따라 자연어 콘텐츠의 전달 루트를 분석하여 해당 내용의 출처 경위를 확인할 수 있다. 계층적 출처 분석부(33)는 자연어 콘텐츠를 출처 경위에 따라 계층적으로 분류하여 지식 기반 데이터베이스로 저장할 수 있다.

[0099] 지식 기반 데이터베이스 구성부(30)는 지식 기반 데이터베이스로부터 질문에 대한 답변 후보군을 추출할 수 있다(S300).

[0100] 지식 기반 데이터베이스 구성부(30)는 지식 기반 데이터베이스에서 질문 입력부(15)에서 입력 받는 질문과 관련된 자연어 콘텐츠를 질문에 대한 답변 후보군으로 추출할 수 있다. 지식 기반 데이터베이스 구성부(30)는 지식 기반 데이터베이스로부터 답변 후보군을 추출하기 위해 질문을 식별 및 분류하여 신뢰성 메트릭을 구성할 수 있다.

[0101] 지식 기반 추론부(50)는 질문 및 답변 후보군을 추론 엔진에 입력하여 답변을 생성할 수 있다(S400).

[0102] 지식 기반 추론부(50)는 질문 문장의 형태에 따라 답변 형태를 참/거진 문답의 답변, 인과 답변 및 확인 답변

중 하나로 설정할 수 있다. 지식 기반 추론부(50)는 답변 형태가 참/거짓 문답의 답변으로 설정되는 경우, 추론 모델로 트리구조 분석 모델을 설정하고, 답변 형태가 인과 답변 또는 확인 답변과 같이 참/거짓 문답의 답변이 아닌 답변으로 설정되는 경우, 추론 모델로 딥러닝 학습 기반의 추론 모델을 설정할 수 있다.

- [0103] 트리구조 분석 모델 질문 및 답변 후보군의 트리구조에 대한 개체명 및 구문론적 비교 분석을 수행하여 질문의 의도에 해당하는 내용이 답변 후보군에 존재하는지 여부에 따라 참/거짓의 답변을 생성할 수 있다.
- [0104] 딥러닝 학습 기반 추론 모델은 답변 후보군의 문장 구조를 분석하여 산출되는 벡터 값을 BERT 딥러닝 네트워크에 입력하여 질문에 대한 답변을 생성할 수 있다.
- [0105] 답변 검증부(70)는 지식 기반 추론부(50)에서 생성한 답변을 검증하고, 출력부(90)는 이를 출력할 수 있다 (S500)
- [0106] 답변 검증부(70)는 지식 기반 추론부(50)에서 생성하는 답변의 신뢰성을 검증할 수 있다. 예를 들면, 답변 검증부(70)는 이전에 사실로 판별된 문장, 사실로 알려진 문장 등에 답변이 있는지 역 추적하여 답변을 검증할 수 있다. 또는, 답변 검증부(70)는 문장을 출처 경위에 따라 분류하여 저장하는 지식 기반 데이터베이스를 이용하여 답변과 다른 출처에 답변과 비슷한 문장이 있는지를 확인하는 방식으로 답변을 검증할 수 있다. 또는, 답변 검증부(70)는 신뢰성 메트릭을 이용하여 답변을 검증할 수 있다. 신뢰성 메트릭은 시간 분류, 출처 식별, 경로 분석, 대명사 출처, 지식 수준 분석 등의 식별 및 분류 작업이 이루어질 수 있다. 답변 검증부(70)는 신뢰성 메트릭의 시간 분류, 출처 식별, 경로 분석, 대명사 출처, 지식 수준 분석 각각에 해당하는 식별 및 분류에 가중치를 더하여, 답변에 해당하는 신뢰도를 수치적으로 계산할 수도 있다.
- [0108] 이와 같은 본 발명의 일 실시예에 따른 지식 기반 추론 및 신뢰도 분석을 이용한 가짜뉴스 판별 방법은 어플리케이션으로 구현되거나 다양한 컴퓨터 구성요소를 통하여 수행될 수 있는 프로그램 명령어의 형태로 구현되어 컴퓨터 판독 가능한 기록 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능한 기록 매체는 프로그램 명령어, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다.
- [0109] 상기 컴퓨터 판독 가능한 기록 매체에 기록되는 프로그램 명령어는 본 발명을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 분야의 당업자에게 공지되어 사용 가능한 것일 수도 있다.
- [0110] 컴퓨터 판독 가능한 기록 매체의 예에는, 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체, CD-ROM, DVD 와 같은 광기록 매체, 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 ROM, RAM, 플래시 메모리 등과 같은 프로그램 명령어를 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다.
- [0111] 프로그램 명령어의 예에는, 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드도 포함된다. 상기 하드웨어 장치는 본 발명에 따른 처리를 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.
- [0113] 이상에서는 실시예들을 참조하여 설명하였지만, 해당 기술 분야의 숙련된 당업자는 하기의 특허 청구범위에 기재된 본 발명의 사상 및 영역으로부터 벗어나지 않는 범위 내에서 본 발명을 다양하게 수정 및 변경시킬 수 있음을 이해할 수 있을 것이다.

부호의 설명

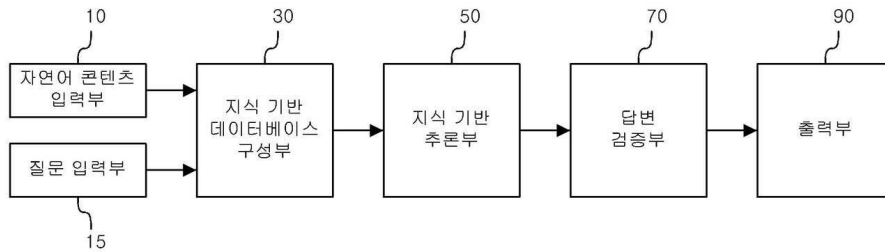
- [0114] 1: 지식 기반 추론 및 신뢰도 분석을 이용한 가짜뉴스 판별 장치
- 10: 자연어 콘텐츠 입력부
- 15: 질문 입력부
- 30: 지식 기반 데이터베이스 구성부
- 50: 지식 기반 추론부
- 70: 답변 검증부

90: 출력부

도면

도면1

1



도면2

30



도면3

50



도면4

