

문제정의서 (연구개발계획서)

프로젝트명	정치, 경제, 시사 분야 가짜 뉴스 판별 모델 개발
-------	------------------------------

조	02 조
지도교수	정상근 교수님 (서명)
조원	201803851 양성욱 201802074 김재운 202104524 박규수

1. 연구개발의 필요성

-가짜 뉴스 판별, 제재, 방지를 위해 현재 국내에서는 2017년부터 컴퓨터 분야 뿐만이 아닌 다양한 분야에서 연구, 논문, 학술지 등이 나오고 있습니다. 이중 저희의 연구와 관련된 컴퓨터 분야에서는 학습 데이터 구축 문제, 자료의 가치중립성, 공정성, 신뢰성 확보, 기계학습 모델의 공신력 문제 등의 문제가 있었고, 문제 대상자를 가짜 뉴스를 필터링하는 관리자를 대상으로 삼아 빠르게 많이 가짜 뉴스들을 걸러내는 것을 목표로 삼고 있었습니다. 이와 연관된 연구들로 국내에서는 한국보건사회연구원에서 “기계학습 기반 보건복지분야 가짜뉴스(fake news) 탐지 방법 연구” 연구 보고서가 있고, 국외에서는 최근 FOX NEWS에서 VERYFIY라는 ai기반 툴을 공개했다. 허나 연구 보고서에서는 가치 중립성이나 공신력등의 문제점을 내세웠고 VERYFIY툴은 모든 사이트에 대한 뉴스를 파악할 수 있는 것은 아니었습니다. 이렇듯 관리자를 대상으로 가짜 뉴스판별을 일괄적으로 하는 방식은 여러 문제점을 보였습니다.

-현재의 대한민국 같이 언론의 신뢰성이 하락하고 가짜 뉴스가 양산되는 탈 진실의 시대에서 가짜 뉴스가 가지는 영향력이 커져만 감에 따라 개인이 제대로 된 근거를 가진 의견을 가지지 못하게 되고 있습니다. 이는 특히 개인의 의견과 참여가 중요한 정치와 시사 분야에서 많은 사람들을 정치 무관심자 혹은 정치 혐오자가 되게 하였으며 개인의 확증편향을 키우고 극단화시키는 문제를 가지고, 정보를 통한 선택의 중요도가 큰 경제 분야에서도 이익을 얻기 위한 가짜 정보로 개인에서 금전적 피해를 주고 있습니다.

-이에 저희의 연구는 위와 같은 상황에서 가짜 뉴스를 판별하지 못하는 정치, 경제, 시사 관련 뉴스기사를 접하는 개인이 가짜 뉴스를 판별하고 근거를 파악하여 자신의 의견을 가지도록 도와 그 과정에 드는 시간을 감축시키고 정확도를 높여 위와 같은 탈 진실 시대에 발생하는 피해를 막고자 합니다.

2. 연구개발의 목표 및 내용

◦ 연구개발의 목표 및 내용

저희의 연구는 기존의 연구들처럼 대량의 가짜 뉴스들을 빠르게 분류하는 관리자를 대상으로 하는 것이 아닌 뉴스기사를 읽고 그 내용을 제대로 파악하고 행동이나 의견을 가지는 것을 목적으로 하는 고관심 계층을 대상으로 합니다. 그렇기에 저희는 여러 뉴스들을 정확하고 빠르게 분류해내는 것보다 대상자의 고충인 가짜 뉴스로 인해 제대로 된 정보를 얻고 판단하기 어려워 진다는 부분을 해결하고자 합니다.

대상자인 개인이 원하는 것은 근거가 제대로된 정보를 얻는 것이므로 저희는 다음과 같은 방식을 생각해 보았습니다.

먼저 딥러닝 모델 학습을 통해 입력받은 뉴스가 내용의 일관성이 없거나, 정보 전달외에 다른 목적이 있는 것 같은 뉴스와 얼마나 비슷한지 백분율로써 나타내줍니다. 그 다음 뉴스 내용을 분석하여 뉴스의 전체 주장, 문단별 요약, 근거 파악에 도움이 되는 문장들, 사용된 문자의 빈도수와 같이 개인이 뉴스를 분석할 때 도움이 될 수 있는 정보를 제공하고, 마지막으로 댓글창을 이용하여 다른 사람의 의견을 보고 추천을 하거나, 자신이 직접 자신의 의견과 근거를 작성하여 공개할 수 있게할 것입니다.

◦ TO-BE 계획

- 1.뉴스 기사를 읽는데 걸리는 시간 : 30분
- 2.뉴스 기사를 읽고 추가로 찾는 자료들 : 5가지 이상
- 3.타인의 의견 접근성 : 3명 이상의 근거가 있는 의견 접근

◦ GAP의 원인

- 1.뉴스 구독자가 기사를 읽으면서 근거에 대한 진위판단 등을 하지 않고 내용만 읽으면서 걸리는 시간이 짧아지고 기사에 대해 생각하지 않음. 또한 진위판단 등을 하면 걸리는 시간의 길이가 매우 길어짐
 - 2.근거에 대한 진위판단이 없으니 해당하는 뉴스기사만을 읽고 추가적인 조사를 하지 않음
 - 3.뉴스 기사가 작성되는 플랫폼은 기본적으로 댓글창이 없거나, 있더라도 자신의 판단에 대한 근거가 없는 경우가 많음
-

3. 연구개발의 추진전략 및 방법

정보 신뢰도 평가 도구 개발: 뉴스 내용의 일관성과 목적을 판별하는 딥러닝 모델을 개발하여, 사용자에게 뉴스의 신뢰도를 백분율로 제공합니다. 이를 통해 사용자는 뉴스의 신뢰도를 더욱 명확하게 인식할 수 있습니다.

뉴스 분석 및 요약 기능 강화: 뉴스 기사의 전체 주장을 추출하고, 문단별 요약을 제공하여 사용자가 빠르고 효과적으로 뉴스 내용을 파악할 수 있도록 합니다. 또한, 근거 파악에 도움이 되는 핵심 문장 및 문자의 빈도수를 제공하여 뉴스 분석을 지원합니다.

사용자 참여 및 의견 공유 기능 강화: 댓글창을 통해 다른 사람의 의견을 확인하고 추천하는 기능을 통하여, 사용자가 직접 의견과 근거를 작성하여 공유할 수 있는 기능을 구현합니다. 이를 통해 다양한 시각과 의견을 공유하고 논의할 수 있는 공간을 제공합니다.

사용자 맞춤형 서비스 개발: 사용자의 관심사와 선호도를 고려하여 맞춤형 뉴스 추천 기능을 제공합니다. 개인화된 서비스를 통해 사용자들이 더욱 효율적으로 원하는 정보를 습득할 수 있도록 합니다.

4. 연구개발 팀의 구성 및 프로젝트 추진 일정

팀 구성	구분	성명	참여내용 및 역할			
	팀장	양성욱	자연어 처리 및 모델 제작			
	팀원	김재운	데이터 셋 수집 및 전처리			
	팀원	박규수	웹 제작 및 DB구축			
추진 일정	세부내용		수행기간(월)			
			3	4	5	6
	데이터셋 수집					
	데이터셋 전처리					
	자연어 처리					
	DB 구축					
	웹 구축					
	모델 제작					

자연어 처리는 뉴스 본문의 주요 내용을 파악하고 키워드를 파악하기 위해 필수적으로 사용된다.

전처리 과정을 통해 뉴스의 키워드를 가져오고 해당 뉴스의 유형을 파악한다.

웹은 해당 모델의 기능을 사용하기위해 링크를 첨부하는 기능, 이전의 파악을 위해 저장한 뉴스의 결과를 저장하기 위해 DB와 웹의 구축이 필요하다. 또한 추가적으로 토론 기능을 더 할 경우 네트워크의 구축도 필요하다.

해당 프로젝트를 진행하기 위해 데이터셋의 전처리와 키워드 수집이 가장 많은 시간을 소모할 것으로 확인된다

- 참고문헌(References)

- 송상현. "가짜뉴스에 대한 규범적·기술적 규제방안 연구." 국내박사학위논문 성균관대학교 일반대학원, 2020. 서울
- 오미애등. (2022). "기계학습 기반 보건복지분야 가짜뉴스(fake news)탐지 방법 연구". 한국보건사회연구원
-<https://www.verify.fox/> (VERIFY TOOL-FOX NEWS)
- 이동호 외6명.(2018.5). "딥러닝 기법을 이용한 가짜뉴스 탐지". 2018년도 춘계학술발표대회 논문집 제25권 제1호(pp.384~387). 한국정보처리학회