



국민대학교
전자정보통신대학
컴퓨터공학부


캡스톤 디자인 I

종합설계 프로젝트

프로젝트 명	팩트체크 프로젝트
팀 명	<i>The Best Plan Is No Plan</i>
문서 제목	계획서

Version	1.2
Date	2018-MAR-08

팀원	황 용 환 (조장)
	김 선 규
	오 근 택
	이 상 욱

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명		
	팀 명		
	Confidential Restricted	Version 1.2	20xx-MAR-25

CONFIDENTIALITY/SECURITY WARNING

이 문서에 포함되어 있는 정보는 국민대학교 전자정보통신대학 컴퓨터공학부 및 컴퓨터공학부 개설 교과목 캡스톤 디자인I 수강 학생 중 프로젝트 "**팩트체크 프로젝트**"를 수행하는 팀 "**The Best Plan Is No Plan**"의 팀원들의 자산입니다. 국민대학교 컴퓨터공학부 및 팀 "**The Best Plan Is No Plan**"의 팀원들의 서면 허락없이 사용되거나, 재가공 될 수 없습니다.

문서 정보 / 수정 내역


Filename	계획서-팩트체크.doc
원안작성자	황용환, 김선규, 오근택, 이상욱
수정작업자	황용환, 김선규, 오근택, 이상욱

수정날짜	대표수정자	Revision	추가/수정 항목	내 용
2018-03-05	전 원	1.0	최초 작성	목차 및 개요 초안 작성
2018-03-07	전 원	1.1	내용 수정	추가 자료조사 및 내용 추가
2018-03-08	전 원	1.2	내용 수정	초안 최종 수정

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명		
	팀 명		
	Confidential Restricted	Version 1.2	20xx-MAR-25

목 차

1	개요	4
1.1	프로젝트 개요	4
1.2	추진 배경 및 필요성	4
2	개발 목표 및 내용	8
2.1	목표	8
2.2	연구/개발 내용	8
2.3	개발 결과	10
2.3.1	결과물 목록 및 상세 사양	10
2.3.2	시스템 기능 및 구조	10
2.4	기대효과 및 활용방안	12
3	배경 기술	12
3.1	기술적 요구사항	12
3.2	현실적 제한 요소 및 그 해결 방안	12
3.2.1	하드웨어	12
3.2.2	소프트웨어	12
3.2.3	기타 오류! 책갈피가 정의되지 않았습니다.	
4	프로젝트 팀 구성 및 역할 분담	13
5	프로젝트 비용	13
6	개발 일정 및 자원 관리	14
6.1	개발 일정	14
6.2	일정별 주요 산출물	15
6.3	인력자원 투입계획	16
6.4	비 인적자원 투입계획	17
7	참고 문헌	18

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명		
	팀 명		
	Confidential Restricted	Version 1.2	20xx-MAR-25

1 개요

1.1 프로젝트 개요

온라인 뉴스기사 중 사회면 뉴스를 입력값으로 한다. 이를 자연어처리를 활용하여 전처리 과정을 거쳐 기계 학습을 할 수 있는 데이터로 가공한다. 기계학습을 통해 모델을 생성하고 새로운 기사를 모델에 적용해 기사의 신뢰도를 보여주는 것을 목표로 한다.

기계학습 방식은 토픽모델링기법의 LDA(Latent Dirichlet Allocation)알고리즘을 사용한다.

1.2 추진 배경 및 필요성

2016년 미국 대선에서 도널드 트럼프가 당선되면서 언론과 전문가들은 당선 이유를 분석하는데 많은 공을 들였다. 다양한 원인들 중 페이크 뉴스(가짜 뉴스)가 큰 영향을 미친 것으로 나타났다. 정체불명의 사이트에서 나온 “교황, 트럼프 지지 선언”, “힐러리, ISIS에 무기 팔아 테러에 사용” 등의 기사들이 페이스북, 트위터 등 소셜 네트워크 서비스를 통해 유통되면서 유권자들의 올바른 선택을 방해했다는 분석이다.

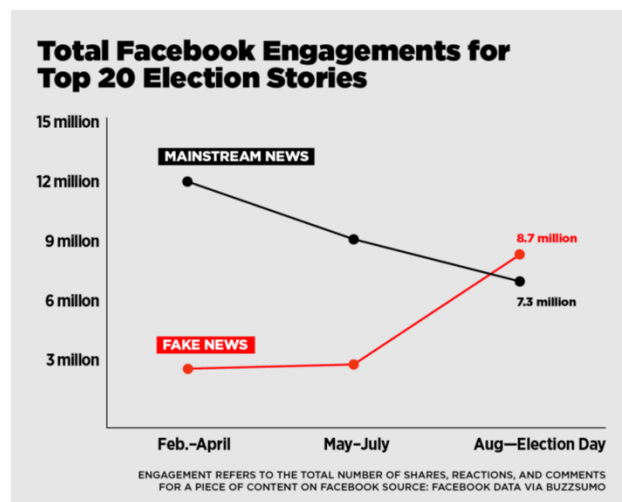



그림 1 페이스북을 통한 진짜뉴스와 가짜뉴스 확산 빈도

출처 : ‘가짜 선거뉴스가 진짜뉴스를 페이스북에서 압도’

<버즈피드>, <https://goo.gl/C8z2Mt>

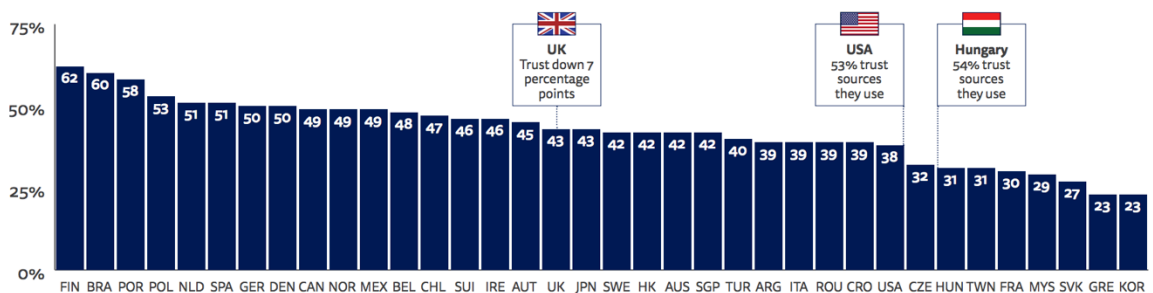
그림1은 지난 미국 대선 때 페이스북을 통한 진짜 뉴스와 가짜뉴스 확산 빈도를 나타낸 그래프이다. 8월 이전에는 진짜뉴스가 가짜뉴스에 비해 3~4배 이상이었지만 8월 이후부터 선거일(2016년 11월 8일)동안의 기간에는 가짜뉴스에 대한 반응(공유, 좋아요 등)이 급등하여 가짜뉴스가 진짜뉴스보다 더 많이 퍼졌다는 것을 확인할 수 있다. 이와 같이 가짜뉴스가 유권자의 올바른 선택에 미치는 영향에 대한 논란이 뜨거웠다. 프랑스, 독일,

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명		
	팀 명		
	Confidential Restricted	Version 1.2	20xx-MAR-25

우리나라 등 선거가 실시되거나 예정되어 있는 나라들에서 가짜 뉴스의 영향을 우려하였고 가짜뉴스는 큰 이슈가 되었다.

뉴욕타임즈에서는 가짜뉴스를 ‘A made up story with an intention do deceive’ 라고 정의했다. 즉, 이야기가 만들어져야 하고 의도적인 속임수가 있어야 가짜뉴스라고 할 수 있다. 기만이나 의도적으로 독자를 혼란시키게 하거나 잘못 이해하게끔 유도하는 뉴스가 가짜뉴스라고 할 수 있다. 가짜뉴스를 생산하는 이유는 미디어 조작 및 선전, 정치 및 사회적 영향, 도발 및 사회적 소요 및 재정적 이익등이 있다. 최근에 일반 뉴스 뿐만 아니라 블로그와 SNS에서 하루에 수 많은 양의 기사들이 뿜어져 나온다. 이러한 언론의 대다수는 광고 기반의 비즈니스 모델로 이동하고 있으며, 기사의 클릭이 주요 경제 동력이며, 그러한 기사는 가짜뉴스 논란의 여지가 있는 경향이 있다. 따라서 뉴스의 사실 검증을 필요로 하게 되었다. 가짜뉴스에 대응하기 위해 무조건적인 규제는 표현의 자유를 침해할 수 있으므로 가짜 뉴스에 대한 대응방안으로 ‘사실 확인(팩트체크)’이 유력한 방안으로 검토되었다.

OVERALL TRUST IN NEWS MEDIA - ALL MARKETS



Q6_2016_1/6. Please indicate your level of agreement with the following statements. - I think you can trust most news most of the time/I think I can trust most of the news I consume most of the time Base: Total sample in each market.

그림 2 국가별 뉴스 신뢰도

출처 : Digital news report 2017


(https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Digital%20News%20Report%202017%20web_0.pdf)

우리나라는 현재 오마이팩트, jtbc팩트체크, 뉴스타파, 뉴스툰, SNU 팩트체크 등 5곳에서 소수인력이 팩트체크를 하여 사실확인 서비스를 제공하고 있다.

하지만 소수 전문화된 인력만으로 사실을 확인하기 때문에 하루에도 생산되는 수많은 기사의 사실을 확인 하기 어려워 일부 사실만을 선택해 확인할 수 밖에 없다. 이 선택 과정에서 사실을 선택하는 것 부터 주관적이기 때문에 편견이 개입될 수 밖에 없다. 따라서 많은 양의 기사를 사실 확인하기 위해 자동화된 사실 확인 기술이 필요로 하게된다.

1.2.1 개발된 시스템 현황

누군가 군 시절 축구경기에서 한 게임에 10골을 넣었다라고 얘기한다면 이를 인간이 사실 검증하는 방법은 크게 4가지가 있다. 첫 번째로는, 군 기록에 남아있는 데이터를 확

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명		
	팀 명		
	Confidential Restricted	Version 1.2	20xx-MAR-25

인하는 것이다. 두 번째로는, 같은 시기에 같은 부대에 근무했던 사람들에게 물어보는 방식이 있고, 세 번째로는 발언한 맥락을 분석한다. 즉 술자리에서 가볍게 하는 이야기 인지, 진지한 분위기에서 서로 진실을 말하는 분위기인지를 파악해야한다. 마지막 네 번째로는 그 사람이 그 동안 얼마나 사실만을 얘기했는지를 알아볼 필요가 있다.

위 4가지 과정을 프로그램으로 적용시킨다면 각각 순서대로 지식기반 방식, 맥락적 방식, 형식 기반 방식, 기계학습 방식이 된다.

구분	대표 사례	장점	단점	한계
지식기반 방식	지식 금고	기록된 데이터를 통한 정확성	방대한 양의 정보 DB화 필요	기록되지 않은 데이터는 확인 불가
맥락적 방식	지식 그래프	기록되지 않은 루머 등도 확인 가능	대중의 잘못된 판단을 사실로 판단할 수 있음	활용하는 소스에 따라 다른 맥락
형식적 방식	‘팩트체크’ 표식	형식 요건을 통해 새로운 사실도 확인 가능	형식만을 판단해 내용에 대해서는 판단하지 못함	형식을 제대로 갖춘 틀린 사실은 확인 불가
기계학습 방식	클레임버스터	사실 여부가 불확실해도 일정 정도 판단 가능	명확한 판단이 아닌 확률적 판단에 불과	모델링의 어려움 및 학습 데이터부족

표 1 자동화된 ‘사실 확인’ 기술의 방식별 사례와 특징

지식기반 방식으로는 DB와 같은 저장소에 사실 내용들을 기사와 페이크, 뉴스인지 아닌지에 대한 값을 라벨링 하여 기록하여서 input이 들어올때 DB와 비교하여서 페이크 뉴스인지 아닌지를 알려준다. 지식기반 방식의 대표적인 예는 지식금고이고, 지식 기반방식의 장점은 기록된 데이터를 통한 정확성이 높다는 것이고, 단점은 방대한 양의 정보 DB화가 필요하다는 것이다. 이 방식의 한계는 기록되지 않은 데이터는 확인이 불가능하다는 것이다.

맥락적 방식은 위키피디아 등 공신력있는 사이트에 검색해서 페이크 뉴스인지 아닌지에 대한 값을 알아내는 것이다. 대표적인 예는 지식 그래프이고, 맥락적 방식의 장점은 기록되지 않은 루머등도 확인이 가능하다는 것이고, 단점은 대중의 잘못된 판단을 사실로 판단할 수 있다. 맥락적 방식의 한계는 활용하는 소스에 따라 맥락이 다를 수 있다.

형식적 방식은 구글에서 사용하는 팩트체크 표식을 남기는 것이다. 이는 구글에서 페이크뉴스가 아닌 기사들의 형식을 정하고 그 형식과 다르다면 페이크 뉴스로 판단하는 방식이다. 이 방식의 장점은 형식 요건을 통해 새로운 사실도 확인 가능하다. 하지만 단점은 형식만을 판단해 내용에 대해서는 판단하지 못한다. 이 방식의 한계는 형식을 제대로 갖춘 틀린 사실은 확인이 불가능하다.


 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명		
	팀 명		
	Confidential Restricted	Version 1.2	20xx-MAR-25

기계학습 방식은 머신러닝을 사용하여 사실여부를 체크할 필요성이 있는 문장을 뽑아내고 사실여부를 판단하는 알고리즘이다. 기계학습 방식의 대표적인 예로는 클레임버스터가 있다. 이 방식의 장점은 사실 여부가 불확실해도 일정 정도 판단 가능하다는 것이고, 단점은 명확한 판단이 아닌 확률적 판단에 불과하다. 기계학습방식의 한계는 모델링의 어려움 및 학습 데이터가 부족하다는 것이다.

1.2.2 개발할 시스템의 필요성

본 프로젝트에서 진행하는 우리의 알고리즘은 입력값으로 기사가 들어온다면 입력값으로 들어오는 기사는 페이크 뉴스가 아니라고 가정한다. 그 기사의 카테고리를 클러스터링하고, 핵심단어들을 추출한다. DB에 저장한 다양한 기사들에게도 카테고리화 핵심언어를 나타내는 TF-IDF값을 저장하여 비슷한 카테고리화 핵심단어가 포함된 유사한 기사들을 뽑아내어 핵심단어가 포함된 문장을 비교한다. 유사한 기사들을 뽑아낼 때 최근의 기사일 수록 중요도를 높게 측정한다. 이 때 페이크 뉴스가 아니라고 가정한 기사와 대립되는 기사들이 많은 경우 해당 기사에 대한 신뢰성이 떨어진다고 판단할 수 있다. 이 방식은 수학적 귀류법과 유사한 알고리즘이다.

위 알고리즘의 장점으로써는 최근 기사에 가중치를 두어 변하는 사실에 대해서도 판단할 수 있다. 단점으로는 루머와 같이 페이크 뉴스이지만 최근에 다량의 기사가 나오는 경우는 판단하기 어렵다. 이 알고리즘의 한계로는 입력값과 유사한 기사가 없을 경우 판단할 수 없다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명		
	팀 명		
	Confidential Restricted	Version 1.2	20xx-MAR-25

2 개발 목표 및 내용

2.1 목표

우리나라에서 페이크 뉴스가 이리이러한 상황에서 퍼지고 있으며 이리이러한 악영향을 끼친다. 어떤 데이터를 활용하여 진위여부, 기사의 신뢰도 또는 검토해야할 부분을 만들어주는 모델을 생성하여 새로운 어떤 데이터를 사실확인 할수 있는 엔진 개발을 목표로 한다.

2.2 연구/개발 내용

- 2.2.1 토픽모델


기계학습 및 자연어처리분야에서 토픽 모델(Topic model)이란 문서 집합의 추상적인 주제를 발견하기 위한 통계적 모델 중 하나로, 텍스트 본문의 숨겨진 의미구조를 발견하기 위해 사용되는 텍스트 마이닝 기법 중 하나이다. 특정 주제에 관한 문헌에서는 그 주제에 관한 단어가 다른 단어들에 비해 더 자주 등장할 것이다. 예를 들어 개에 대한 문서에서는 "개"와 "뽀다귀"라는 단어가 더 자주 등장하는 반면, 고양이에 대한 문서에서는 "고양이"와 "야옹"이 더 자주 등장할 것이고, "그", "~이다"와 같은 단어는 양쪽 모두에서 자주 등장할 것이다. 이렇게 함께 자주 등장하는 단어들은 대개 유사한 의미를 지니게 되는데 이를 잠재적인 주제로 정의할 수 있다. 즉, "개"와 "뽀다귀"를 하나의 주제로 묶고, "고양이"와 "야옹"을 또 다른 주제로 묶는 모형을 구상할 수 있는데 바로 이것이 토픽 모델의 개략적인 개념이다. 실제로 문헌 내에 어떤 주제가 들어있고, 주제 간의 비중이 어떤지는 문헌 집합 내의 단어 통계를 수학적으로 분석함으로써 알아 낼 수 있다.

- 2.2.2 TF-IDF

TF-IDF(Term Frequency - Inverse Document Frequency)는 정보검색과 텍스트 마이닝에서 이용하는 가중치로, 여러 문서로 이루어진 문서군이 있을 때 어떤 단어가 특정 문서 내에서 얼마나 중요한 것인지를 나타내는 통계적 수치이다. 문서의 핵심어를 추출하거나, 검색 엔진에서 검색 결과의 순위를 결정하거나, 문서들 사이의 비슷한 정도를 구하는 등의 용도로 사용할 수 있다.

TF는 단어의 빈도를 나타내며 특정한 단어가 문서 내에 얼마나 자주 등장하는지를 나타내는 값으로, 이 값이 높을수록 핵심단어라고 할 수 있다. 하지만 단어 자체가 문서군 전체 내에서 자주 사용되는 경우, 흔한 단어를 의미하며 이를 DF라고한다. TF값이 높을수록, DF값이 낮을수록 핵심단어이다.

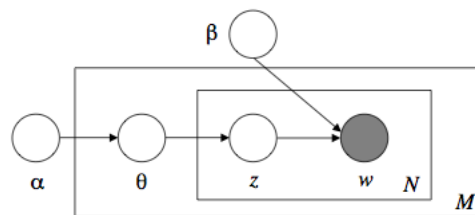
-

 <div> <p>국민대학교</p> <p>컴퓨터공학부</p> <p>캡스톤 디자인 I</p> </div>	계획서		
	프로젝트 명		
	팀 명		
	Confidential Restricted	Version 1.2	20xx-MAR-25


- 2.2.3 LDA알고리즘

LDA(Latent Dirichlet Allocation)는 문서 같은 데이터의 집합에 대한 Generative Probabilistic Model이다. α , β 는 코퍼스 단위로 정해지는 값, N 과 θ 는 문서 단위로 정해지는 값, β 는 각 주제별로 특정 단어가 생성될 확률이 담긴 테이블, N 은 문서의 길이, θ 는 해당 문서에서 각 주제의 가중치를 나타낸다.

어떤 문서에 대해 파라미터 θ (주제 벡터)가 있고, 앞에서부터 단어를 하나씩 채울 때마다 θ 로 부터 하나의 주제를 선택하고, 다시 그 주제로 부터 단어를 선택하는 방식으로 문서 생성 과정을 모델링하는 것이다. 아래 그림은 이 과정을 요약한 것이다.

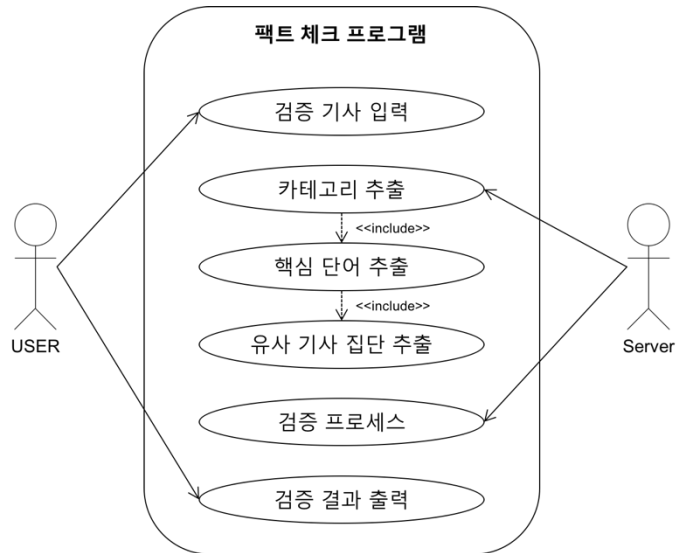


이런 방법으로 문서 내용을 성공적으로 모델링, 즉 표현할 수 있다면 이미 존재하는 문서의 파라미터 θ 를 찾아내는 것도 가능할 것이다. 글 d_1 과 d_2 가 있을 때, 주제는 비슷하더라도 각 문서에 등장하는 단어의 종류나 빈도는 다를 수 있기 때문에 단순한 키워드 기반의 모델로는 유사도를 계산하거나 주제 분류를 하는 데에 한계가 있다. 그러나 이미 보유한 많은 텍스트에 기초에 α 와 β 를 알아 두고, 개별 문서의 θ 를 계산할 수 있다면, 이 θ 를 가지고 유사도 계산이나 분류 작업을 훨씬 쉽고도 정확하게 해낼 수 있다는 것이다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명		
	팀 명		
	Confidential Restricted	Version 1.2	20xx-MAR-25

2.3 개발 결과

2.3.1 시스템 기능 요구사항




사용자가 검증하고 싶은 기사를 입력한다. 서버에 있는 팩트체크 엔진을 이용하여 기사의 카테고리화 핵심단어를 추출한다. 팩트체크 엔진이 카테고리가 유사한 기사 목록들을 추출하고 사실 여부를 검증한다.

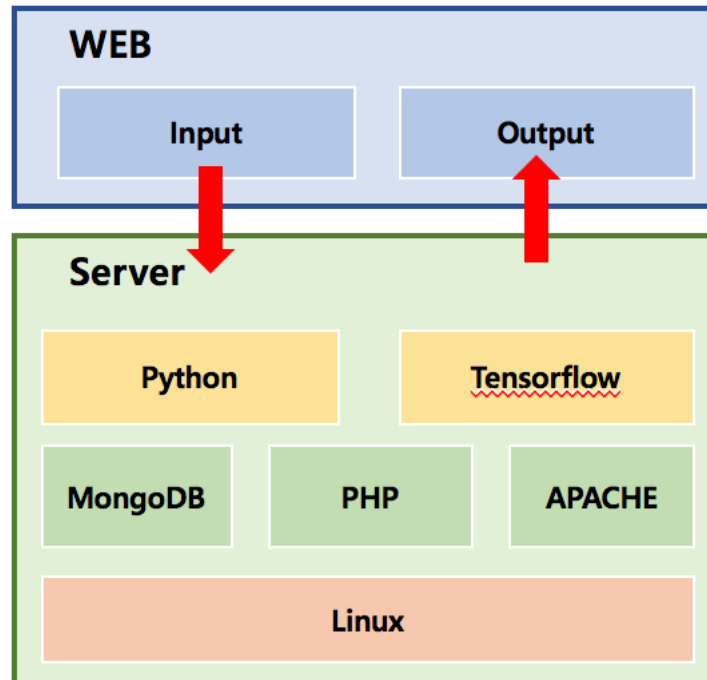
2.3.2 시스템 비기능(품질) 요구사항

1. 사용자가 입력한 기사의 대한 검증 결과를 얻기 위한 응답시간을 30초 이내로 하도록 알고리즘이 수행.
2. 팩트 체크 모델의 학습 데이터를 적어도 1000개 이상을 확보하여서 정확성을 높인다.
3. 사용자가 편리하게 이용하도록 기사 링크와 원문 모두를 지원하는 WEB UI를 설계.(usability)


※ 우선순위는 내림차순이다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명		
	팀 명		
	Confidential Restricted	Version 1.2	20xx-MAR-25

2.3.3 시스템 구조



웹에서 input Data 를 받아서 Server 의 파이썬과 Tensorflow 를 통해서 자연어 전처리 기능을 수행하며 전반적인 팩트체크 시스템이 수행된다. 서버는 리눅스 서버를 사용하며 DB 는 Mongo DB 를 사용한다. 서버에서 생성된 output 은 WEB 으로 보여진다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명		
	팀 명		
	Confidential Restricted	Version 1.2	20xx-MAR-25

2.4 기대효과 및 활용방안

일반적인 관점에서 볼 때 이 기술은 소비자들이 무엇을 믿어야 할 지 결정할 수 있도록 함으로써 정보 소비 습관을 변화시킬 수 있다. 또한 이해를 향상시킬 수 있으며 사회의 양극화를 최소화 할 수 있다.

현재 우리나라에서는 팩트체크에 대해서는 일반적으로 전문가들이 직접 한다. 예를 들어서 오마이팩트, JTBC 팩트체크, 뉴스타파, 뉴스룸, SNU 팩트체크 총 5곳이 있다. 하지만 이 5곳은 모두 전문 인력이 직접 팩트체크를 한다. 예를 들어 JTBC 팩트체크에서는 전문 인력이 직접 다양한 기사들을 찾아보며 팩트여부를 직접 판단하게 된다. 이는 사람이 직접하므로 팩트체크에서 가장 중요한 객관성이 흐려질 수 있다.

소비자들은 뉴스 기사 유통의 플랫폼화 때문에 뉴스를 언론사 홈페이지보다 SNS나 유통 플랫폼에서 더 자주 보게 된다. 뉴스 유통 플랫폼에서는 뉴스를 직접 생산하기보다 유통에만 집중을 하기 때문에 사실확인을 할 전문 인력을 보유하기 어렵다.

즉, 팩트체크의 공정성 및 객관성 측면과 전문인력 부족으로 인해 팩트체크 프로그램의 활용도가 기대된다.

3 배경 기술

3.1 기술적 요구사항

서버 운영체제 환경 : Ubuntu 16.04 LTS
 서버 개발 환경 : Django, PHP, Apache
 개발 언어 : Python 3.6 이상
 라이브러리 : KoNLPy(자연어처리), TensorFlow(Machine Learning)

3.2 현실적 제한 요소 및 그 해결 방안

3.2.1 하드웨어

팩트 체크 모델을 학습시키는 속도가 느려 정확성이 떨어 질 수 있다.
 GPU를 사용하여 속도를 늘린다. 혹은 AWS ec2를 사용한다.

3.2.2 소프트웨어

데이터의 양과 질 또는 데이터의 전처리과정에 따라 팩트체크 모델의 정확도가 떨어질 수 있다. 따라서 학습모델을 만들 시 다양한 파라미터를 이용해서 최적값을 구한다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명		
	팀 명		
	Confidential Restricted	Version 1.2	20xx-MAR-25

4 프로젝트 팀 구성 및 역할 분담

이름	역할
황용환	<ul style="list-style-type: none"> - Software Project Leader - 사전 조사 - 자연어 전처리
김선규	<ul style="list-style-type: none"> - 서버 프로세스 - LDA 알고리즘 최적화
오근택	<ul style="list-style-type: none"> - 머신러닝 최적화 - 크롤링
이상욱	<ul style="list-style-type: none"> - 머신러닝 최적화 - 사전 조사

5 프로젝트 비용

항목	예산치 (MD)
이슈 분석	20
아이디어 구상	40
크롤링	60
자연어 전처리	40
비지도학습 클러스터링 연구	40
LDA 알고리즘	20
서버 구축	20
웹 UI 제작	10
합	250

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명		
	팀 명		
	Confidential Restricted	Version 1.2	20xx-MAR-25

6 개발 일정 및 자원 관리

6.1 개발 일정

항목	세부내용	1 월	2 월	3 월	4 월	5 월	6 월	비고
요구사항분석	요구 분석							
	아이디어 구상							
관련분야연구	자연어 처리 연구							
	머신러닝 연구							
설계	시스템 설계							
구현	크롤링							
	LDA 알고리즘							
	LSTM 알고리즘							
	웹 UI							
테스트	시스템 테스트							

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명		
	팀 명		
	Confidential Restricted	Version 1.2	20xx-MAR-25

6.2 일정별 주요 산출물

마일스톤	개요	시작일	종료일
계획서 발표	개발 환경 완성 (Ubuntu 데스크탑 2 대 설치) 산출물 : 1. 프로젝트 수행 계획서 2. PPT 및 영상	~	2018-03-09
중간 보고	크롤링 및 데이터 클러스터링 및 LDA 알고리즘 완료 산출물 : 1. 프로젝트 중간 보고서 2. 프로젝트 진도 점검표 3. 중간 구현 소스 코드	2018-03-09	2018-04-13
구현 완료	팩트체크 구현 완료 산출물: 소스코드	2018-04-13	2018-05-19
테스트	시스템 통합 테스트 산출물: 테스트 결과 보고서	2018-05-19	2018-05-26
최종 보고서	최종 보고 산출물: 최종보고서	2018-05-26	2018-05-29

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명		
	팀 명		
	Confidential Restricted	Version 1.2	20xx-MAR-25


6.3 인력자원 투입계획

이름	개발항목	시작일	종료일	총개발일(MD)
전 원	이슈 분석 및 아이디어 구상	2018-01-02	2018-03-09	60
전 원	크롤링	2018-03-08	2018-04-01	60
황 용 환	자연어 전처리	2018-03-15	2018-04-07	40
이 상 욱 오 근 택	비지도학습 클러스터링 연구, LDA 알고리즘	2018-04-12	2018-04-20	40
김 선 규 황 용 환	LSTM 알고리즘	2018-04-12	2018-04-20	20
김 선 규 이 상 욱	서버구축	2018-03-28	2018-04-17	20
전 원	웹 UI 제작	2018-04-28	2018-05-19	10

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명		
	팀 명		
	Confidential Restricted	Version 1.2	20xx-MAR-25

6.4 비 인적자원 투입계획

항목	Provider	시작일	종료일	Required Options
서버용 데스크탑	SAMSUNG	2018-01-10	2018-05-29	
개인 노트북 4 대	Lenova, Apple	2018-01-10	2018-05-29	

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명		
	팀 명		
	Confidential Restricted	Version 1.2	20xx-MAR-25

7 참고 문헌

번호	종류	제목	출처	발행년도	저자	기 타
1	논문	Discriminative Predicate Path Mining for Fact Checking in Knowledge Graphs	RISS	2016	Baoxu Shi and Tim Weninger	
2	논문	Fact Checking in Heterogeneous Information Networks	RISS	2017	Baoxu Shi Tim Weninger	
3	논문	Finding Streams in Knowledge Graphs to Support Fact Checking	RISS	2016	Prashant Shiralkar	
4	논문	Toward Computational Fact-Checking	RISS	2016	You Wu	
5	Web page	Detecting Fake News with Scikit-Learn	DataCamp			
6	Web page	How can Machine Learning and AI help solving the Fake news Problem?	miguelmalvarez			
7	기사	가짜뉴스 찾는 ‘인공지능 R&D 챌린지’본선 개최	ZD NET Korea			
8	기사	가짜뉴스 꼼짝 마! 팩트체크 인공지능 나선다	동아사이언스			
9	Web page	Fake News Challenga	FNC			
10	Web page	SNU fact check	서울대학교 언론정보연구소			