# FakeText:FastText 기반의 한국어 가짜 뉴스 분류 모델

강효은, 심준석, 김용수, 홍윤영, 김호원*

부산대학교

# *FakeText:FastText-based Korean Fake News Classification Model*

Hyoeun Kang, Junseok Shim, Yongsu Kim, Yoonyoung Hong, Howon Kim*

Pusan National University

**요 약**

최근 주식시장의 열기가 뜨거워짐에 따라 주식시장의 변동성을 예측할 수 있는 뉴스의 영향력도 커지고 있다. 동시에 기업가치를 왜곡하는 가짜 정보 확산이 사회적으로 큰 영향을 미치고 있다. 해외의 경우 가짜 뉴스 유포 문제를 해결하기 위해 구글과 페이스북 등 유명 IT기업들에서 솔루션을 제안하고 있으며, 국내의 경우에도 텍스트 마이닝, 머신러닝/딥러닝을 활용한 가짜뉴스 판별에 대한 연구가 활발히 진행 중이다. 본 논문은 한국어 가짜 뉴스를 판별하기 위해 형태소 단위로 학습된 워드 임베딩 모델인 FastText과 딥러닝 모델을 활용한 한국어 가짜뉴스 판별 모델을 제안한다.

## I. 서론

가짜뉴스는 정치 및 경제적 이익을 위해 의도적으로 언론 보도의 형식을 한 거짓 정보이다. 가짜뉴스는 의도적으로 만들어지고 오해를 부를 수 있는 허위정보, 고의로 조작한 정보를 사실을 가장하는 거짓정보, 근거없이 퍼지는 소문인 루머 등의 형태로 퍼지기도 한다[1].

2016년 미국 45대 대선 과정에서 페이크 뉴스가 광범위하게 파급되어 선거에 영향을 미쳤으며[1], 미국뿐만 아닌 우리나라에서도 최근 코로나19 재확산에 따른 방역 당국이 정치적인 목적으로 진단 검사 결과를 조작하고 있다는 가짜뉴스 사례가 유포되는 등 사회적으로 큰 영향을 미치고 있다. 최근 국내 주식시장의 열기가 뜨거워지면서 주가조작을 위한 가짜뉴스도 확산되고 있다. 주로 뉴스를 통해 기업의 내재가치를 분석하는 투자자들은 언제든지 주가조작 피해 위험에 노출될 수 있다.

가짜뉴스의 유형은 기사 제목과 본문이 부정합한 경우와 본문 중 맥락에 관계가 없는 내용으로 나눌 수 있다. 본 연구에서는 본문의 흐름에서 벗어나는 왜곡된 문장을 탐지하여 가짜뉴스 여부를 판별하는 모델을 제안한다.

본 논문에서는 성능이 우수하다고 알려진 공개용 한국어 형태소 분석기 중 하나인 Mecab-ko[7]와 FastText[5]을 이용하여 뉴스 데이터의 언어 표현을 학습한다. 최종적으로 딥러닝 모델을 이용하여 가짜뉴스 여부 판별을 진행한다.

### 1.2 관련 연구

구글은 가짜 리뷰 탐지를 위해 SVM을 사용해 리뷰어의 행동 데이터와 리뷰를 결합하여 분석하는 방법을 소개하였으며[2], 페이스북의 FiB[8]는 자바스크립트 기반으로 링크, 포스트, 이미지 정보를 추출하고 이를 딥러닝을 통해 기사의 원출처를 확인하여 가짜뉴스를 판별한다.

국내의 경우 Yun Tae-Uk[3]는 토픽 모델

링 기법을 사용하여 가짜뉴스를 분류하다. Ye-Chan Ahn[4]는 한국어 BERT 사전 학습 모델에 Fine tuning을 적용한 가짜뉴스를 판별한다.

1.3 워드 임베딩

워드 임베딩(Word embedding)이란 단어를 벡터로 표현하는 방법으로, 단어를 희소 표현과 밀집 표현으로 구분된다. 희소 표현에는 대표적으로 표현하고자 하는 단어의 인덱스의 값만 1로 표현하고 나머지 인덱스를 0으로 표현하는 원-핫 벡터 표현 방법이 있다. 해당 방법을 사용하면 단어의 개수가 늘어남에 따라 벡터의 차원이 한없이 커진다는 단점이 있다.

한편 밀집 표현에 해당하는 방식에는 대표적으로 FastText[5]가 있다. FastText는 주변 단어와 단어의 부분 단어(subword)를 학습하여 학습 말뭉치에 없는 Out of Vocabulary(OOV)에 해당하는 단어의 임베딩 값을 얻을 수 있다. 본 연구에서 사용하는 딥러닝 모델은 FastText의 임베딩 결과 값을 입력으로 받아 가짜뉴스 여부를 판별한다.

## II. 데이터 전처리

2.1 데이터세트

본 연구는 실제로 국내에서 이슈가 되었던 가짜뉴스를 판별하는 모형을 만들기 위해 다수의 선행 연구에서 활용되었던 SNU Factcheck와 뉴스톱, 팩트체크넷 사이트에서 총 2,066개의 데이터세트를 구축하였다. 진짜뉴스는 1,056개, 가짜뉴스는 1,010개를 수집하여 실험 데이터세트를 완성하였다. 실험을 위한 데이터 분할은 Train 1,446개, Validation 310개, Test 310개를 사용하였다.

2.2 데이터 전처리

본 연구에서는 형태소 분석을 이용해 어휘에서 의미적(semantic)인 부분이 아닌 것을 제거하였다. 형태소 분석에는 오픈 한국어 형태소 분석기인 Mecab-ko을 사용하였다.

## III. FastText 기반 가짜뉴스 판별 모델

3.1 FastText

한국어는 교착어에 속하기 때문에 다양한 접사들이 하나의 어근에서 약 60여 가지의 단어로 파생될 수 있다[6]. 한국어와 같은 교착어는 어절 단위로 그대로 학습할 경우 개별 단어의 빈도가 대부분 낮기 때문에 유의미한 정보를 얻기 어렵다. 본 연구에서는 n-gram 알고리즘을 적용하여 형태학적(morphological) 특성을 벡터 값에 반영할 수 있는 FastText를 사용하였다. 오탈자 및 오류가 거의 없는 한국어 위키피디아 문서를 기반으로 2백만개의 단어에 대하여 300차원으로 한국어 어휘 임베딩을 학습한 사전학습 모델을 사용하였다.

3.2 지도학습 기반 가짜뉴스 판별 딥러닝 모델

본 연구에서는 본문의 맥락에서 벗어나는 문장에 대하여 가짜뉴스 판별을 하기 위해 랜덤포레스트, CNN, Bidirectional LSTM 3가지 분류 모델을 구현하였다.

## IV. 실험 결과

본 논문에서 제안한 가짜뉴스 판별 모델의 성능 검증을 위해 표 1과 같이 정확도와 FPR, FNR을 사용하였다. FPR은 정상 뉴스를 가짜뉴스로 오분류한 비율을 나타내며 FNR은 가짜뉴스를 정상 뉴스로 판단한 비율을 나타낸다. FastText와 BiLSTM을 결합한 모델이 우수한 성능을 보임을 확인할 수 있다.

| 모델 / 성능 | 랜덤포레스트 | CNN | **BiLSTM** |
|---|---|---|---|
| Acc (%) | 92.80 | 97.10 | **98.90** |
| FPR | 0.1245 | 0.0390 | **0.0160** |
| FNR | 0.3317 | 0.0544 | **0.0213** |

표 1. 가짜뉴스 판별 모델 평가 결과

## V. 결론

본 연구는 FastText 워드임베딩 모델을 이용하여 한국어 뉴스의 어휘 특징을 추출하고, 딥러닝 모델을 적용하여 가짜뉴스 여부를 판별하였다. 본문이 전반적으로 정상적인 정보를 전달하고 일부 문장이 거짓 정보를 의미하는 데이

터에 대하여 가짜뉴스 판별 모델을 설계하였다. 추후 본 연구에서 구현한 모델을 사용하여 구독자 또는 정보 수용자들이 뉴스의 맥락을 통해 사실 관계를 합리적으로 이해할 수 있는 플랫폼으로 확장하여 구현하고자 한다.

## 감사의 글

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터지원사업의 연구결과로 수행되었음 (IITP-2021-0-01797), 블록체인 기반 및 플랫폼 분야 핵심기술 개발 및 미래 혁신인재 양성

## [참고문헌]

[1] 윤성옥. 가짜뉴스의 개념과 범위에 관한 논의. 언론과법, 2018, 17.1: 51-84.

[2] MUKHERJEE, Arjun, et al. Fake review detection: Classification and analysis of real and pseudo reviews. UIC-CS-03-2013. Technical Report, 2013.

[3] YUN, Tae-Uk; AHN, Hyunchul. Fake News Detection for Korean News Using Text Mining and Machine Learning Techniques. Journal of Information Technology Applications and Management, 2018, 25.1: 19-32.

[4] AHN, Ye-Chan; JEONG, Chang-Sung. Natural language contents evaluation system for detecting fake news using deep learning. In: 2019 16th International Joint Conference on Computer Science and Software Engineering (JCSSE). IEEE, 2019. p. 289-292.

[5] BOJANOWSKI, Piotr, et al. Enriching word vectors with subword information. Transactions of the Association for Computational Linguistics, 2017, 5: 135-146.

[6] 김한샘.현대국어사용빈도조사.Vol.2.국립국어원, 2005

[7] 은전한닢 Mecab-ko, https://bitbucket.org/eunjeon/mecab-ko/src/master/

[8] FiB, “FiB: Lets stop living a lie,” 2016.