ISSN 2288-4866 (Print)
ISSN 2288-4882 (Online)
https://www.jiisonline.org
J Intell Inform Syst 2023 June: 29(2): 267~283 http://dx.doi.org/10.13088/jiis.2023.29.2.267

품사별 출현 빈도를 활용한 코로나19 관련 한국어 가짜뉴스 탐지*

김지혁
국민대학교 비즈니스IT전문대학원
(kjh9654@kookmin.ac.kr)

안현철
국민대학교 비즈니스IT전문대학원
(hcahn@kookmin.ac.kr)

2019년 12월부터 현재까지 지속되고 있는 코로나19 팬데믹으로 인해 대중들은 감염병 대응을 위한 정보를 필요로 하게 되었다. 하지만 소셜미디어에서 유포되는 코로나19 관련 가짜뉴스로 인해 대중들의 건강이 심각하게 위협받고 있다. 특히 코로나19와 관련된 가짜뉴스가 유사한 내용으로 대량 유포될 경우 사실인지 거짓인지 진위를 가리기 위한 검증에 소요되는 시간이 길어지게 되어 우리 사회의 전반에 심각한 위협이 될 수 있다. 이에 학계에서는 신속하게 코로나19 관련 가짜뉴스를 탐지할 수 있는 지능형 모델에 대한 연구를 활발하게 수행해 오고 있으나, 대부분의 기존 연구에 사용된 데이터는 영문으로 구성되어 있어 한국어 가짜뉴스 탐지에 대한 연구는 매우 드문 실정이다. 이에 본 연구에서는 소셜 미디어 상에서 유포되는 한국어로 작성된 코로나19 관련 가짜뉴스 데이터를 직접 수집하고, 이를 기반으로 한 지능형 가짜뉴스 탐지 모델을 제안한다. 본 연구의 제안모델은 언어학적 특성 중 하나인 품사별 빈도 정보를 추가적으로 활용하여, 기존 연구에서 주로 사용되어 온 문서 임베딩 기법인 Doc2Vec 기반 가짜뉴스 탐지 모델의 예측 성능을 제고하고자 하였다. 실증분석 결과, 제안모델이 비교 모델에 비해 Recall 및 F1 점수가 높아져 코로나19 관련 한국어 가짜뉴스를 보다 정확하게 판별함을 확인하였다.

주제어 : 코로나19 관련 가짜뉴스, 한국어 가짜뉴스, 소셜 미디어, Doc2Vec, 품사 구분

논문접수일 : 2023년 5월 15일 논문수정일 : 2023년 6월 12일 게재확정일 : 2023년 6월 19일
원고유형 : 학술대회 Fast Track 교신저자 : 안현철

1. 서론

가짜뉴스(fake news)는 대중들로 하여금 잘못된 정보로 현혹하고 사회적으로 혼란과 불안이 발생하도록 유도하여 심각한 문제로 대두되고 있다. Vosoughi et al.(2018)는 2013년 버락 오바마 전 미국 대통령이 폭발 사고로 부상을 입었다는 허위 트윗(tweet)으로 인해 1,300억 달러의 주식 가치가 사라졌던 사례와 자연재해, 테러 등 인류에게 위협이 되는 사건들에 대한 대응이 가짜뉴스로 인해 차질을 빚은 적이 있었다는 사례를 근거로 가짜뉴스가 경제적, 사회적으로 큰 파장을 미치고 있다고 주장하였다. Bovet et al.(2019)은 2016년 미국 대통령 선거를 예로 들며, 실제 뉴스 보도를 심각하게 왜곡하는 가짜뉴스는 특유의 자극적이고 참신함을 무기로 대중들을 매료시켜 마치 바이러스처럼 소셜 미디어 네트워크(SNS)에서 전파된다고 주장하여 가짜뉴스의 심각성을 강조하였다.

그런 와중에 2019년 12월에 처음 시작되어 현재까지도 지속되고 있는 코로나19 팬데믹(COVID-19

* 이 논문 또는 저서는 2022년 대한민국 교육부와 한국연구재단의 인문사회분야 중견연구자지원사업의 지원을 받아 수행된 연구임(NRF-2022S1A5A2A01048638)

pandemic)으로 인해 인류는 감염병 대응을 위한 정보를 절실히 필요로 하게 되었다. 코로나19와 관련된 유익한 정보가 다양한 대중매체를 통해 전해졌으나, 동시에 대중을 현혹시키기 위한 가짜뉴스도 범람해 이른바 인포데믹(infodemic)이 발생하였다. 인포데믹은 신뢰할 수 있는 매체의 역할을 저해하며, 인류의 건강에 큰 위협이 되고 있다(한지원, 김영욱, 2023). 이종구(2020)는 코로나19 방역을 위해 집 안에 메탄올을 뿌렸다가 일가족이 중독 증세를 보여 병원에 입원한 사례와 어느 교회에서 예배를 드리기 전 신도들에게 분무기로 소금물을 뿌린 사례 등 실제로 국내에서 발생했던 사례들을 언급하며 인포데믹으로 인한 정보의 오류와 위험성을 강조하면서, 신중하고 정확한 정보에 대한 필요성을 주장했다.

국립재난안전연구원(2020)은 뉴스, SNS, 개인 동영상 플랫폼 등에서 1995년에서 2020년까지 발생했던 인포데믹의 발생 확률과 원인을 조사하였다. 조사 결과 지진과 태풍 같은 자연재난보다 감염병이나 미세먼지와 같은 사회적 재난이 발생했을 때 인포데믹의 발생 확률이 가장 높았던 것으로 나타났다. 보고서에서는 이러한 현상의 원인으로 학자마다 발생여부 및 시기, 장소, 원인 등에 대해 다른 견해를 가지고 있어 뉴스에 대한 불확실성이 높아진다는 점을 꼽았다. 또한 부정확하고 편향된 허위 정보들이 대량으로 유포되면 전문가들도 사실인지 거짓인지 구분하기 어려워져 검증에 소요되는 시간이 길어진다는 점도 함께 지적하였다. 이러한 인포데믹 문제에 대한 심각성을 인식하고 이를 예방하기 위해 SNU 팩트체크나 팩트체크넷과 같은 플랫폼들이 등장하면서 다양한 분야의 전문가들과 시민들이 함께 정보의 진실 여부를 판별하고 있다. 그러나 이러한 플랫폼들은 결국 사람이 수동으로 검증한다는 한계 때문에 많은 시간과 비용이 소모되며, 예산 삭감 등의 이유로 인해 문을 닫는 플랫폼도 등장하고 있는 실정이다(박수선, 2023).

이에 학계에서는 가짜뉴스를 자동으로 탐지하는 인공지능 기반의 지능형 모델을 연구하여 가짜뉴스를 근절하기 위한 노력을 기울이고 있다. 하지만 이러한 지능형 모델 연구가 수행되기 위해서는 충분한 양의 가짜뉴스 학습 데이터가 요구되는데, 코로나19 관련 가짜뉴스의 경우 최근에 유통이 시작되었기 때문에 축적된 양이 많지 않고, 진위여부 검증이 쉽지 않아 공개된 학습용 데이터셋이 매우 드물다. 그럼에도 불구하고 선도적으로 코로나19 관련 가짜뉴스 데이터셋 구축 및 탐지 연구를 시도한 초기 연구들을 살펴보면, Patwa et al.(2021)과 한소은 등(2021)은 트위터에서 코로나19와 관련된 가짜뉴스를 수집하여 데이터셋을 구축하였고, 이를 활용하여 Ngada & Haskins(2020)와 심재승 등(2020)이 가짜뉴스를 탐지하기 위한 기법을 제시하였다. 이지민 외(2022)는 각종 매체에서 유통되는 코로나19와 관련된 뉴스의 신뢰성을 판단하기 위해 가짜뉴스 판별 웹페이지를 구축하기도 하였다. 하지만 국내에서 진행된 코로나19 관련 가짜뉴스 탐지 연구들은 대부분 영문으로 작성된 해외 데이터셋을 사용하고 있다는 한계가 있었다.

이에 본 연구에서는 코로나19 관련 한국어 가짜뉴스 탐지 연구를 위해서는 데이터 확보가 선행되어야 한다는 점을 인식하고, 대표적인 소셜미디어 중 하나인 트위터에서 유포되고 있는 한국어로 작성된 코로나19 관련 가짜뉴스를 수집하여 이를 기반으로 공개 데이터셋을 구축하였다. 이어 해당 데이터셋을 기반으로 한국어에 특화된 코로나19 관련 가짜뉴스 탐지 모델을 개발하고자 하였다. 특히 본 연구에서는 영화 리뷰의

감성 분석에서 품사 정보를 활용하는 것이 정확도 향상에 기여하는 것으로 나타난 정세민 등(2021)의 연구 결과를 참고하여 품사별 빈도 정보를 추가로 활용함으로써 가짜뉴스 탐지 성능 제고를 꾀하였으며, 실증분석을 통해 실제로 성능이 개선되는지를 확인하고자 하였다.

이후 논문의 구성은 다음과 같다. 2장에서는 본 연구에서 적용될 가짜뉴스의 정의와 지금까지 연구된 데이터셋과 탐지 모델에 대한 설명, 그리고 제안 모델의 핵심이 되는 임베딩(embedding) 기법인 Word2Vec 및 Doc2Vec의 원리를 소개한다. 이어 3장에서는 제안 모델의 구조와 함께 실험을 위한 데이터 전처리 과정에 대해서 설명한다. 4장에서는 실증분석을 통해 제안하는 가짜뉴스 탐지 모델과 기존의 텍스트 기반 가짜뉴스 탐지 모델의 성능을 비교한다. 마지막으로 5장에서는 본 연구의 결과를 정리하고, 본 연구의 학술적 의의, 한계점 및 후속 연구 방향에 대해 소개한다.

2. 이론적 배경

2.1. 가짜뉴스의 정의에 관한 연구

가짜뉴스에 대한 정의는 아직까지 합의된 것이 없기 때문에, 학자마다 가짜뉴스를 바라보는 시각에 따라 조금씩 차이가 있다(이장근 등, 2022). Lazer et al.(2018)는 가짜뉴스를 형식상 뉴스 미디어 콘텐츠와 유사한 형태를 보이지만 조작된 정보로서 거짓 혹은 오해의 소지가 있는 정보를 의도적으로 유포하여 사람들을 속이기 위해 사용되는 것으로 정의하였으며, Tandoc et al.(2018)은 가짜뉴스가 단순히 뉴스의 패러디 혹은 정치적인 풍자와 같은 것이 아니라 허위 기사를 유포하는 것과 동시에 일부 언론사의 비판적 보도를 폄하하는 의도로도 사용되어 그 경계가 희미하다고 주장하였다. 민희(2022)는 이전과는 달리 가짜뉴스라는 개념이 보편화 되었기 때문에 기사 형식으로 작성한 조작된 정보뿐만 아니라 소셜 네트워크에서 전파되는 찌라시 혹은 낚시성 기사, 전체 기사 중 일부분만 짜깁기한 편파적인 기사, 클릭수를 높이기 위해 반복 게재하는 기사 등도 가짜뉴스의 범주로 취급해야 한다고 주장했다. 이에 본 연구에서도 민희(2022)의 주장을 수용하여 넓은 범주로 정의해야 한다고 판단했으며, 이에 따라 가짜뉴스를 “조작되고 잘못된 정보, 낚시성 기사 및 광고 목적이 강한 기사, 그리고 기사의 일부를 짜깁기하여 진실을 왜곡한 뉴스”로 정의하였다.

2.2. 코로나19 관련 가짜뉴스 데이터셋 연구

Patwa et al.(2021)은 트위터에서 코로나19와 관련된 게시물을 수집하고 해당 콘텐츠가 진짜인지 가짜인지 라벨을 붙여 구분한 코로나19 관련 가짜뉴스 데이터셋을 제시하였다. 이 데이터셋은 Politifact와 Newschecker, Boomlive와 같은 다양한 팩트 체크 사이트에서 검증된 기사를 바탕으로 구성하였다. 수집된 데이터는 신뢰할 수 있는 출처에서 코로나19에 대한 유용한 정보를 제공한 경우 ‘Real’로 라벨링을 수행하고, 잘못된 정보나 주장 및 추측이 포함된 경우 ‘Fake’로 라벨링을 수행하였다. 수집된 데이터는 모두 코로나19와 관련된 콘텐츠이며, 사용된 언어는 오직 영어로 한정했다. 자체적으로 각종 기계학습 모델을 통해 가짜뉴스 판별 실험을 수행한 결과, SVM에서 Accuracy 93.24%, Precision 93.48%, Recall 93.46%, F1 점수 93.46%로 높은 성능을 보였다.

한편 Cui & Lee(2020)는 CoAID라는 명칭의

가짜뉴스 데이터셋을 개발하여 코로나19 가짜뉴스 탐지를 수행하였다. 이 데이터셋은 2019년 12월부터 2020년 7월까지 유통된 코로나19 관련 뉴스로 구성된 영문 데이터셋이다. 해당 데이터셋은 진짜뉴스 3,055개와 가짜뉴스 866개로 구성된 총 3,921개의 뉴스 데이터와 제목 및 URL을 참조한 193,312개의 트윗 정보로 구성되었다. 이후 한소은 등(2022)은 CoAID 데이터셋에 각 트윗의 좋아요 수, 리트윗 수 등 소셜 활동 정보를 추가하여 발전시켰다. 그리하여 전통적으로 사용되던 내용 기반의 가짜뉴스 탐지 모델 개발은 물론이고 전파 정보를 기반으로 한 소셜 컨텍스트 기반의 가짜뉴스 탐지 모델 개발에도 적용이 가능한 CoAID+를 제안하였다. 이처럼 코로나19와 관련된 가짜뉴스 탐지를 위해 다양한 데이터셋 연구가 진행되었으나, 이 데이터셋들은 모두 영어로 구성된 데이터셋이라는 점에서 한계가 있다.

2.3. 기존 가짜뉴스 탐지 연구

Bondielli & Marcellon(2019)에 의하면 가짜뉴스를 탐지하는 방법은 내용 정보 기반 기법(content based methods)과 소셜 컨텍스트 기반 기법(social context based methods)으로 나눌 수 있다. 이 중 내용 정보 기반 기법은 뉴스의 내용에서 언어적 특징을 찾아서 가짜뉴스를 탐지하는 방법이다. 반면에 소셜 컨텍스트 기반 기법은 사용자 정보, 사용자 간 관계, 좋아요(like)나 리트윗(retweet)과 같은 공유 행동 등 이용자들의 행동 정보를 활용하여 가짜뉴스를 탐지하는 방법이다.

2.3.1. 소셜 컨텍스트 기반 가짜뉴스 탐지 기법

현윤진과 김남규(2018)는 뉴스 데이터에서 추출한 주제적 특성과 해당 뉴스에 대한 트위터 내 반응에 대한 벡터 값을 결합한 가짜뉴스 탐지 모델을 제안했다. 이 모델은 뉴스 데이터를 토픽모델링을 통해 구조화한 후 클러스터링을 진행하여 이슈 그룹을 생성하였다. 이후 각 뉴스에 속한 각각의 트윗에 대해 벡터 값을 추출하는 방법으로 뉴스 데이터의 속성과 트위터 데이터의 속성을 같이 고려한 결과, 높은 정확도를 보이는 것이 확인되었다.

정이태와 안현철(2022)은 사회적 참여 네트워크 정보를 활용하여 코로나19 가짜뉴스를 탐지하는 소셜 컨텍스트 기반 탐지 모델을 제안했다. 이 모델은 네트워크 상에서 발생하는 참여 관계를 그래프의 형태로 가공하여 사회적 참여 네트워크를 사용하였다. 본 연구는 앞서 소개한 CoAID 데이터셋을 사용했으며, 뉴스 데이터의 제목은 DistiRoberta(A distilled version of BERT)를 활용하여 임베딩하고 가공된 그래프를 Graph2vec을 통해 특징 변수로 활용한 그래프 임베딩 기반의 탐지모델을 제안하였다.

2.3.2. 내용 기반 가짜뉴스 탐지

Ngada & Haskins(2020)는 가짜뉴스 텍스트와 그 언어학적 특성을 고려한 가짜뉴스 탐지 모델을 제안하였다. 이들은 Kaggle에서 배포한 2015년~2018년 동안의 가짜뉴스 데이터셋을 사용하여 Doc2Vec 모델로 워드 임베딩을 수행하고, 기사 제목 및 본문의 총 단어 수와 총 부호 수, 기사 내에 있는 품사 태그의 수, 평균 문장 길이 등 기사의 내용과 구조를 설명하는 특징 변수를 채택하였다. 그런 다음 SVM, Decision Tree, Random Forest, KNN, XGB의 기계학습 분류기를 사용하여 가짜뉴스 탐지 모델을 성능을 평가하였다. 그 결과 단순히 문서의 벡터값만 활용했을 경우보다 기사 본문의 구조의 특징을 포함한 변수를 추가했을 때 훨씬 좋은 정확도를 제공한다는 것을 확인하였다.

심재승 등(2020)은 Word2Vec을 활용해 특징을 추출한 가짜뉴스 탐지 모델을 제안하고, SNU 팩트 체크와 네이버 뉴스를 통해 2,188건의 한국어 가짜뉴스 데이터셋을 구축하여 제안모델의 성능을 검증하고자 하였다. 구체적으로 인공신경망(Artificial Neural Network)을 이용하여 Word2Vec으로 특징을 추출했을 경우와 TF-IDF로 특징을 추출했을 경우를 비교하였는데, 실증분석 결과 Word2Vec을 적용한 제안모델이 비교모델보다 월등히 정확도를 개선한다는 점을 확인하였다. 이 연구는 워드 임베딩을 실제 한국어 가짜뉴스 탐지에 적용한 첫 번째 연구라는 점에서 학술적 의의가 있다.

2.4. 워드 임베딩

2.4.1. Word2Vec

Mikolov et al.(2013)에 의해 제안된 Word2Vec은 원래 상태로는 분석이 불가능한 자연어를 컴퓨터가 이해할 수 있도록 숫자의 형태로 임베딩하는 기법이다. Word2Vec는 유사한 의미의 단어들은 서로 가까운 거리를 가지며, 유사한 문맥에서 함께 등장한다고 전제한다. 이러한 전제를 바탕으로 주어진 문장을 구성하는 단어들의 관계를 파악하고 그 의미를 내포한 벡터(vector) 값으로 수치화하는 Word2vec은 문장의 의미를 파악해야 하는 추론에 효과적으로 적용될 수 있다(정예림 등, 2020).

Word2Vec의 학습 방법은 CBOW(Continuous Bag Of Words)와 Skip-gram 중 하나로 결정된다. 이 중, CBOW는 주변 단어를 통해 중심(목표)에 위치한 단어를 예측하는 방법이다. CBOW에서는 창(window)의 크기를 통해 주변의 선별할 단어의 수를 설정하고, 선별한 단어들로 중심단어를 One-hot Vector로 표현한다. 예를 들어 “자신을 화나게 했던 ___을 다른 이에게 행하지 말라” 라는 문장이 있다면 중심 단어 근처에서 좌우로 각각 3개의 단어인 “자신을”, “화나게”, “했던”, “다른”, “이에게”, “행하지”를 통해 중심 단어를 예측하도록 학습시킨다.

반면에 Skip-gram 방법은 CBOW 방법과는 반대로 중심에 위치한 단어를 활용해 주변에 어떤 단어가 등장할 지 예측하는 방법이다. 예를 들면 “___ 화나게 ___”라는 문장이 있다면 “화나게” 라는 단어를 활용해 주변 단어를 예측한다.

2.4.2. Doc2Vec

Doc2Vec은 Le & Mikolov(2014)가 제안한 방법으로서, 문단이나 문서를 임베딩하기 위해 Word2Vec을 발전시킨 기법이다. 즉, Word2Vec이 단어와의 거리를 파악하고 단어에 대해서 벡터 값을 생성하는 방법이었다면, Doc2Vec은 문장과 문단, 문서 단위로 벡터 값을 임베딩한다(송찬우, 안현철, 2022). 구체적으로 Doc2Vec은 각 문맥별로 등장하는 단어들의 분포 특성을 추출하고 이를 결합하는 방법으로서, 문장 전체에서 k개의 단어가 주어질 때 그 다음에 등장할 단어를 예측하는 과정을 통해 학습을 진행한다. Doc2Vec은 이 과정에서 로그 확률 평균을 최대로 만들도록 학습을 진행하는데, 이 값이 높아질수록 훈련 시간은 길어진다는 단점이 있으나 정확도가 높아질 수 있다. 로그 확률 평균의 계산 방법은 다음과 같다.

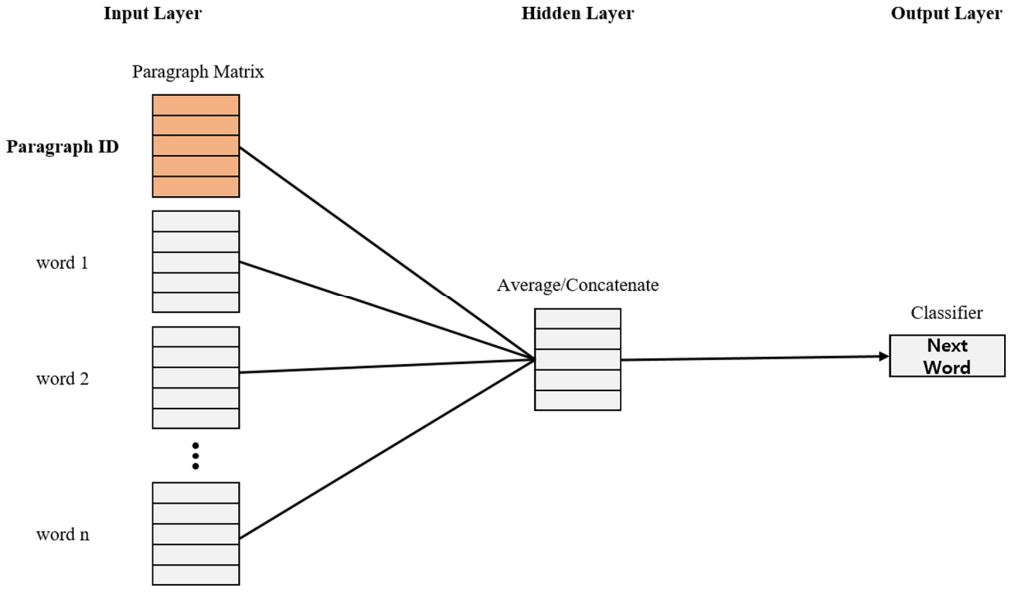
$$\frac{1}{T}\sum_{t=k}^{T-k}\log p(w_t|w_{t-k},\dots,w_{t+k})$$

w_t: 문장 내 t번째 단어

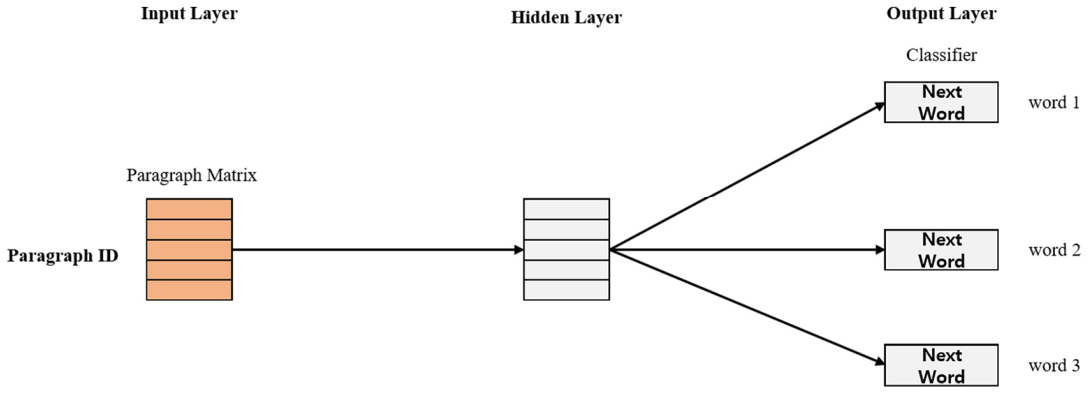
Doc2Vec을 학습시키는 방법으로는 PV-DM과 PV-DBOW가 있다. PV-DM(Distributed Memory Model of Paragraph Vectors)는 Word2Vec의 CBOW

방법과 유사한 학습 방법으로, <그림 1>과 같이 k개의 단어에 문서ID(paragraph ID)를 활용해 다음 단어를 예측한다. 이 때 문서ID가 함께 학습되기 때문에 임베딩이 해당 문서의 주제 정보를 포함할 수 있게 되었고, 단어가 등장하는 순서도 고려되기 때문에 기존의 Bag-of-Words 방법보다 발전했다고 볼 수 있다.

반면 PV-DBOW(Distributed Bag of Words version of Paragraph Vector)는 <그림 2>와 같이 문서ID만 입력해 일정 수의 단어를 예측하며 학습하는 방법이다. 오직 문서ID만을 사용하기 때문에 단어 벡터를 고려하지 않고 단어 예측을 수행하며, 단어들이 순서와 상관없이 랜덤으로 추출된다는 특징이 있다.



〈그림 1〉 PV-DM 방법



〈그림 2〉 PV-DBOW 방법

3. 연구모델

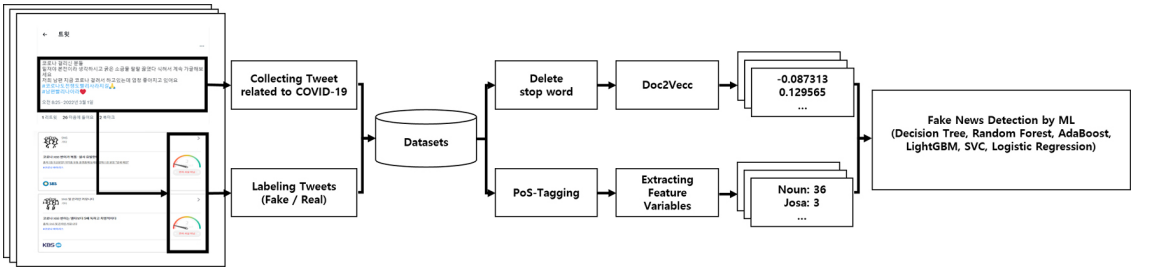
본 연구에서 제안하는 코로나19 관련 한국어 가짜뉴스 탐지 모델의 전반적인 구조는 다음의 <그림 3>과 같다.

본 연구에서는 우선 각종 코로나19 관련 데이터셋과 팩트체크 사이트를 참고하여 사전에 가짜뉴스로 판별된 뉴스들의 주요 키워드를 활용하여 트위터에서 비슷한 내용의 트윗을 수집한다. 이어 수집된 트윗에 Doc2Vec을 적용해 임베딩을 진행한 벡터값과 POS Tagging을 거친 품사의 등장 빈도를 특징변수로 사용하여 한국어로 작성된 코로나19 가짜뉴스를 탐지하는 모델을 구축한다. 이후 각종 기계학습 분류 알고리즘을 활용해 결과를 비교한다. 즉, 본 연구의 제안모델은 트윗의 텍스트를 임베딩한 결과와 품사의 출현 빈도라는 2종의 특징변수들을 활용하며, 각각의 전처리 과정은 다음과 같다.

3.1. 품사의 출현 빈도

본 연구에서는 텍스트 형태의 트윗별 품사의 등장 빈도를 추출하기 위해 Okt(Open-Korean-Text) 형태소 분석기를 채택하였다. Okt[1]는 트위터에서 개발한 Twitter 형태소 분석기를 기반으로 탄생한 한국어에 특화된 형태소 분석기이며, 짧은 문장의 트윗은 물론 긴 글의 형태소 분석 역시 가능하다. 김수연 등(2022)에 의하면 Okt가 Kkma나 Komoran에 비해 POS Tagging의 정확도가 다소 떨어진다는 단점이 있으나, 표현이 다른 단어를 같은 단어로 통합해 주는 정규화 기능을 활용할 수 있다는 점에서 유리하다. 예를 들어 '저는 학교에 가고 있습니닿ㅎㅎㅎ' 이라는 문장은 정규화를 통해 '저는 학교에 가고 있습니다ㅎㅎ'로 수정할 수 있으며, POS Tagging을 수행하면 문장 마지막의 'ㅎㅎ'가 KoreanParticle이라는 별도의 태그로 분류된다.

원혜진 등(2020)에 의하면 Okt는 Hannanum이나 Komoran과는 다르게 자체적으로 띄어쓰기가 되어 있지 않은 문장에 자동으로 공백을 삽입하여 전처리에 효율적이며, 다른 형태소 분석기보다 신조어를 잘 분류한다는 장점을 가지고 있다. 또한 Okt는 Hashtag, URL 그리고 KoreanParticle을 포함한 21개의 독자적인 품사 태그를 정의하고 있다. 안형준(2020)에 의하면 어떤 주제에 대해 적극적인 정보 공유가 많이 발생하는 경우 이용자들이 Hashtag 및 URL의 사용 빈도가 증가하게 되는 SNS의 언어심리학적 특징이 있다고 언급하고 있어, 이러한 SNS의 특성도 반영할 수 있을 것으로 판단된다. 이에 본 연구에서는 Okt를 형태소 분석기로 채택하였다.



〈그림 3〉 제안 모델의 구조

1) https://github.com/open-korean-text/open-korean-text

Okt를 활용하여 각 트윗별로 품사 빈도를 추출할 때, 별도의 불용어 처리는 진행하지 않았다. 그리고 품사 빈도를 추출할 때 명사와 같이 많이 등장할 수 있는 품사가 있다는 점을 고려해 각 품사 빈도별로 최소-최대 정규화(Min-Max Scaler)를 수행하였다. 각 트윗의 품사별 빈도 정보를 추출하는 예시가 아래의 <표 1>에 제시되어 있다.

3.2. 텍스트 임베딩

본 연구에서는 텍스트의 특징 정보를 추출하기 위해 앞서 언급한 Doc2Vec을 사용해 텍스트 임베딩을 진행하였다. 임베딩을 진행하기 전에 앞서 트위터에서 가져온 코로나19 관련 정보를 포함한 트윗 텍스트에 길호현(2018)이 제시한 불용어 대상 어휘와 '코로나19', 'COVID-19' 등과 같이 자주 등장하는 단어를 대상으로 불용어 처리를 수행하였다. 해당 과정 이후 Doc2Vec을 이용해 수집한 트윗들을 벡터화하였다. 비교적 짧은 길이의 텍스트인 트윗을 대상으로 데이터를 수집했기 때문에 Vector Size는 30으로 작게 설정하였다. 그 외 파라미터로 window는 5, min_alpha는 0.0001, epochs는 50을 적용했고, min_count를 5로 설정해 단어의 빈도가 5번 이하 등장할 경우 무시하도록 하였다.

4. 실증 분석

4.1. 실험데이터

본 연구에서 사용한 코로나19 관련 한국어 가짜뉴스 데이터는 대표적인 소셜미디어 서비스 중 하나인 트위터를 이용해 수집하였다. 트위터는 자신이 주변 지인들에게 공유하고 싶은 일상을 짧은 인스턴트 메시지(instant message) 형태로 공유하는 마이크로블로그(micro-blog)의 형태를 띄고 있어서, 지금도 많은 이용자들이 사용하고 있다는 장점이 있다(심홍진, 황유선, 2010). 또한 카카오톡이나 네이버 밴드 등의 폐쇄형 플랫폼은 아예 데이터 수집을 금지하고 있고, 페이스북 역시 특정 그룹에서의 자료만 수집할 수 있는 한계가 있는데, 트위터는 공개적으로 데이터를 수집할 수 있어서 연구에 활용하기에 유리하다(손승혜 등, 2018). 게다가 트위터는 리트윗(retweet) 기능을 제공하여 가짜뉴스 확산에 최적인 소셜 미디어라는 특징도 갖고 있어, 본 연구에서는 트위터를 중심으로 코로나19 관련 가짜뉴스 데이터를 수집하는 것이 가장 적절할 것으로 판단하였다. 본 연구에서 트위터를 통해 자료를 수집한 과정을 상세히 설명하면 다음과 같다.

우선 트위터에서 데이터로 사용할 트윗을 검색

〈표 1〉 트윗별 품사 빈도 정보 추출 예

Tweet	CDC에서 코로나 바이러스 대비 마스크 쓸 때 수염 있는 사람들은 주의하라고 가이드라인 냈는데 수염 종류가 이렇게 많을 줄이야 ㅋㅋㅋㅋㅋㅋㅋ
Tokenization	[('CDC', 'Alpha'), ('에서', 'Josa'), ('코로나', 'Noun'), ('바이러스', 'Noun'), ('대비', 'Noun'), ('마스크', 'Noun'), ('쓸', 'Verb'), ('때', 'Noun'), ('수염', 'Noun'), ('있는', 'Adjective'), ('사람', 'Noun'), ('들', 'Suffix'), ('은', 'Josa'), ('주의', 'Noun'), ('하라고', 'Verb'), ('가이드', 'Noun'), ('라인', 'Noun'), ('냈는데', 'Verb'), ('수염', 'Noun'), ('종류', 'Noun'), ('가', 'Josa'), ('이렇게', 'Adverb'), ('많을', 'Adjective'), ('줄', 'Noun'), ('이야', 'Josa'), ('ㅋㅋㅋㅋㅋㅋㅋ', 'KoreanParticle')]
Feature Variable	'Alpha':1, 'Noun' : 10, 'Josa':4, 'Verb':2, 'Adjective':2, 'Suffix':1, 'Korean Particle':1
Feature Variable (Min_Max Scaler)	'Alpha': 0.076923, 'Noun' : 0.234043, 'Josa': 0.333333, 'Verb': 0.153846, 'Adjective': 0.222222, 'Suffix': 0.666667, 'Korean Particle': 0.333333

〈그림 4〉 진짜뉴스 워드 클라우드

〈그림 5〉 가짜뉴스 워드 클라우드

한다. 수집한 데이터셋의 신뢰성을 높이기 위해 사전에 가짜뉴스인지 검증된 데이터로 공공데이터 포털에서 제공하는 '한국언론진흥재단_뉴스빅데이터_메타데이터_코로나' 데이터셋[2]과 '한국언론진흥재단_뉴스빅데이터_메타데이터_가짜뉴스' 데이터셋[3]을 참고하여 국내 언론사에서 실제로 배포한 코로나19 관련 뉴스의 키워드를 추출하고, 트위터에서 검색을 통해 유사한 기사가 있었는지 검색하여 정리한다. 국제적으로 유행했던 코로나19 가짜뉴스는 Patwa et al.(2021)의 데이터셋을 참고하여 검색에 활용한다. 이렇게 정리된 한국언론진흥재단의 데이터셋은 다시 뉴스데이터베이스인 'BIGKinds'로 검색하여 코로나19와 관련된 뉴스의 게시일자, 언론사, 기고자, 제목, 키워드, 본문을 수집한다. 진짜/가짜뉴스의 라벨링과 관련해서는, 사전에 가짜뉴스로 판별된 내용을 포함한 트윗과 트윗의 내용과 전혀 무관한 사이트로 연결되는 URL이 포함된 트윗, 기사의 일부분만 오묘하게 짜깁기한 트윗은 가짜(fake)로 라벨링한다. 반면에 보건복지부 및 질병관리청과 같이 신뢰할 수 있는 기관에서 유포한 트윗 및 이를 토대로 작성된 트윗은 진짜(real)로 라벨링한다.

데이터 불균형을 방지하기 위해 가짜뉴스 트윗과 진짜뉴스 트윗은 각각 157건씩 1대 1비율로 수집하였다. 이렇게 최종 정리된 본 연구에 사용된 데이터는 다른 연구자들도 참고하여 활용할 수 있도록 Github[4]에 공개하였다. 다음의 <그림 4>와 <그림 5>는 각각 트위터에서 유통된 코로나19 관련 한국어 진짜뉴스와 가짜뉴스의 단어별 출현 빈도를 시각화한 워드 클라우드(word cloud)를 나타내고 있다. 이 워드 클라우드들로 미루어 볼 때, 가짜뉴스의 경우 주로 약물이나 성분과 관련된 단어가 상대적으로 많이 출현하고 있음을 알 수 있다.

이렇게 수집된 데이터에 대해 앞서 3.1과 3.2에서 설명한 방법으로 전처리를 진행하였다. 품사별 빈도 정보를 파악하기 위해 불용어 처리를 진행하지 않고 Okt를 사용하여 특징 변수를 생성하였

2) https://www.data.go.kr/data/15086437/fileData.do
3) https://www.data.go.kr/data/15069309/fileData.do
4) https://github.com/KISLABatKMU/COVID-19_Korean_Fake_News_in_Twitter

으며, 트윗의 내용을 벡터화하기 위해 Doc2Vec을 이용해 문서 임베딩을 수행하여 30차원의 벡터로 전처리를 진행하였다. 이어서 사이킷런(sklearn) 라이브러리의 Train_Test_Split으로 학습용 70%, 시험용 30%로 데이터셋을 구분하였으며, 동일한 데이터셋을 사용하여 실험을 진행하기 위해 Random_State는 42로 고정하였다.

4.2. 실험 모델

본 연구에서는 코로나19 관련 한국어 가짜뉴스 탐지를 위한 분류 기법으로 Decision Tree, Random Forest, SVC(Support Vector Classification), Logistic Regression과 AdaBoost, CatBoost의 Boosting 계열 기법들을 적용하였다. 이 중, Decision Tree는 데이터셋의 크기가 작고 단순하기 때문에 과대적합 방지를 위해 max_depth를 2로 제한을 두고, min_samples_leaf는 5로 설정하였으며, Random Forest는 min_samples_split을 10으로 설정하였다. SVC의 경우에는 C를 2, gamma를 5로 설정하였고, Logistic Regression 역시 C를 2로 설정하였다. AdaBoost는 n_estimators를 100, learning_rate를 0.1로 설정하였으며, CatBoost는 iterations을 100, learning_rate를 0.1, depth를 2로 설정하고 실험을 진행하였다.

본 실험에서의 비교모델은 Doc2Vec으로 텍스트를 벡터화하여 그 결과만 사용하는 모델로 설정하였으며, 제안모델은 Doc2Vec 결과에 더해 품사별 출현 빈도를 추가 특징 변수로 활용한 모델로 설정하였다. 그렇게 하여 이 2가지 모델을 트위터 상에서 유통되는 코로나19 관련 한국어 가짜뉴스 탐지에 적용했을 때, 제안모델이 비교모델 대비 더 우월한 성과를 보이는지 확인해 보고자 하였다.

4.3. 실험 결과

본 연구에서는 가짜뉴스 탐지 모델 평가를 위해 시험용 데이터셋의 Accuracy와 학습용 데이터셋의 Accuracy를 각각 계산하고 이 둘의 너무 차이가 벌어지지 않도록 하여 과적합을 방지하였다. 그리고 기준이 되는 모델의 평가지표로는 시험용 데이터셋에서의 Accuracy, Precision, Recall, 그리고 F1 점수를 사용하였다. 여기에서 Precision은 탐지 모델이 가짜뉴스라고 분류한 것들 중 실제로 가짜뉴스인 경우의 비율을 나타내며, Recall은 실제로 가짜뉴스인 트윗 중 모델이 가짜뉴스로 판별한 비율을 의미한다. F1 점수는 Precision과 Recall의 조화평균으로 산출한다. 각 평가지표들은 다음의 수식을 통해 계산할 수 있다(Shu et al., 2017).

$$\text{Accuracy} = \frac{|TP| + |TN|}{|TP| + |TN| + |FP| + |FN|}$$

$$\text{Precision} = \frac{|TP|}{|TP| + |FP|}$$

$$\text{Recall} = \frac{|TP|}{|TP| + |FN|}$$

$$\text{F1}_{\text{score}} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

TP (True Positive): 가짜뉴스로 예측했으며 실제로도 가짜뉴스였던 경우
TN (True Negative): 진짜뉴스로 예측했으며 실제로도 진짜뉴스였던 경우
FN (False Negative): 진짜뉴스라고 예측했으나 실제로는 가짜뉴스였던 경우
FP (False Positive): 가짜뉴스라고 예측했으나 실제로는 진짜뉴스였던 경우

전반적인 실험 결과는 <표 2> ~ <표 4>에 정리

되어 있다. <표 2>는 텍스트를 벡터화하여 이 결과로 가짜뉴스를 탐지하는 전통적인 가짜뉴스 탐지 모델에 대한 평가 결과이며, <표 3>은 본 연구에서 제안하는 가짜뉴스 탐지 모델로서, 품사의 출현 빈도를 같이 고려한 모델에 대한 평가 결과이다. <표 4>는 제안 모델을 채택했을 경우, 비교 모델보다 어느 정도 성능지표의 향상이 있었는지 정리한 표이다.

〈표 2〉 벡터화된 텍스트만으로 가짜뉴스를 탐지한 결과: 비교모델 결과

Model	Accuracy	Precision	Recall	F1 점수
Decision Tree	0.811	0.875	0.778	0.824
Random Forest	0.853	0.900	0.833	0.865
SVC	0.832	0.852	0.852	0.852
Logistic Regression	0.779	0.902	0.685	0.779
AdaBoost	0.779	0.824	0.778	0.800
CatBoost	0.863	0.918	0.833	0.874

〈표 3〉 품사별 등장 빈도 정보를 추가로 활용해 가짜뉴스를 탐지한 결과: 제안모델 결과

Model	Accuracy	Precision	Recall	F1 점수
Decision Tree	0.863	0.873	0.889	0.881
Random Forest	0.884	0.922	0.870	0.895
SVC	0.853	0.885	0.852	0.868
Logistic Regression	0.842	0.898	0.815	0.854
AdaBoost	0.821	0.863	0.815	0.838
CatBoost	0.884	0.906	0.889	0.897

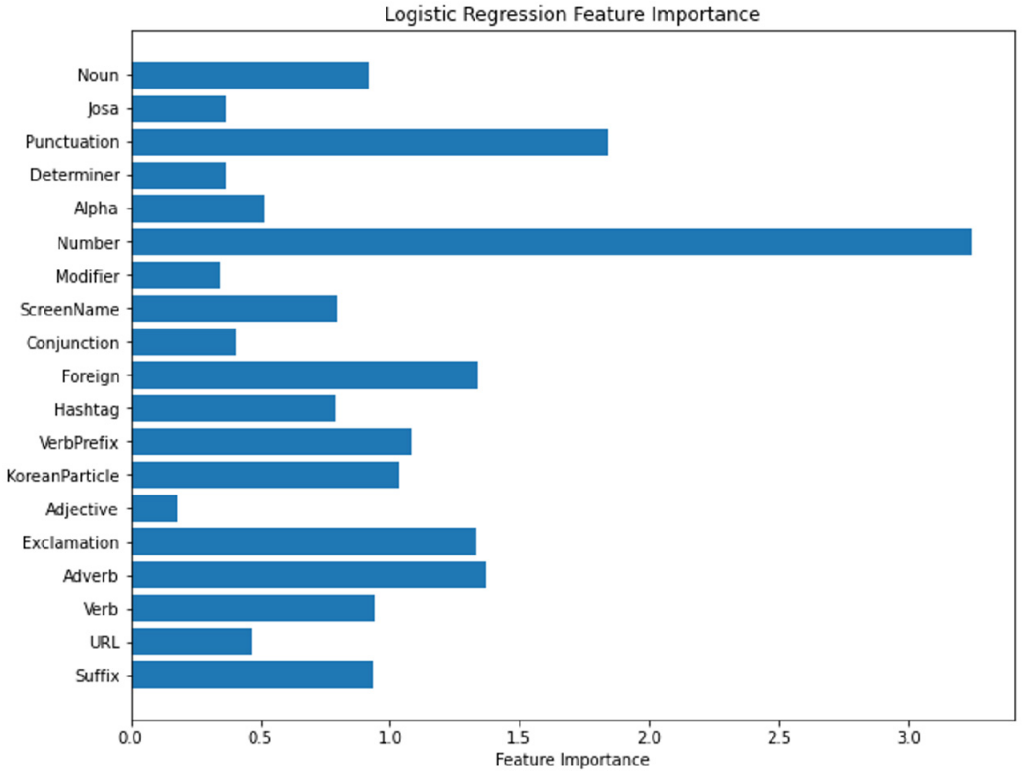
〈표 4〉 제안모델과 비교모델 간 성과 차이

Model	Accuracy	Precision	Recall	F1 점수
Decision Tree	0.025	-0.002	0.111	0.057
Random Forest	0.031	0.022	0.037	0.030
SVC	0.021	0.033	0.000	0.016
Logistic Regression	0.063	-0.004	0.130	0.075
AdaBoost	0.042	0.039	0.037	0.038
CatBoost	0.021	-0.012	0.056	0.023

전체적으로 보았을 때, Decision Tree와 Logistic Regression의 Precision이 미세하게 감소했고, SVC의 Recall 값이 변동이 없었다는 점을 제외하면 모든 경우에서 탐지 모델의 성능이 향상되었음을 확인할 수 있다. 품사별 출현 빈도를 추가로 고려할 경우, Logistic Regression의 F1 점수가 7.5%로 가장 높은 상승폭을 보였고, 이어서 Decision Tree 5.7%, AdaBoost 3.8%, Random Forest 3%, CatBoost 2.3%, SVC 1.6% 순으로 성능의 향상을 보였다. 앞서 언급했던 대로 Decision Tree와 Logistic Regression의 Precision은 결과 표를 비교해 볼 때 약간의 감소가 발생했으나, 전체적인 Accuracy는 0.052, 0.063 증가하였기 때문에 종합적인 관점에서는 품사별 출현 빈도를 사용하는 것이 코로나19 관련 가짜뉴스 탐지 성능을 제고하는데 뚜렷하게 기여하고 있음을 확인할 수 있다. 본 연구에서 적용한 여러 기법들 중에서 F1 점수의 상승폭이 가장 높았던 Logistic Regression에서 나타난 품사별 특성 중요도는 다음의 <그림 6>과 같다. 이 결과에 따르면 특히 'Number(수치)'와 'Punctuation(문장부호)'이 높은 특성 중요도를 보임을 확인할 수 있다. 진짜뉴스가 정부 기관의 공식 SNS 계정 등에서 문서의 마침표 사용에 준수하는 형식을 갖고 있으며 명확한 수치를 내용에 포함하여 게시되는 반면, 가짜뉴스는 문서의 형식적인 제약이 없으며 자극적인 내용을 구성하기 위해 일부 키워드만을 사용하여 전파되는 경향이 있는데, 이러한 특성이 반영된 결과인 것으로 해석된다.

5. 결론

본 연구에서는 지난 코로나19 팬데믹 기간 동안 대중들을 현혹하고 우리 사회를 위험에 노출



〈그림 6〉 Logistic Regression으로 도출된 각 품사별 중요도

시켰던 코로나19 관련 가짜뉴스를 판별하는 지능형 탐지 모델을 제안하였다. 본 연구는 특히 한국어로 된 데이터셋이 부재하여 한국 상황에 특화된 코로나19 관련 가짜뉴스 탐지 연구가 수행되기 어려운 환경적 제약을 극복하고자, 모델 개발에 앞서 코로나19 관련 한국어 가짜뉴스 데이터셋 구축을 진행하였다. 구체적으로 연구에 많이 사용되고 있는 Patwa et al.(2021)의 영문 코로나19 가짜뉴스 데이터셋과 SNU팩트체크와 같은 한국의 팩트체크 사이트를 참고한 내용을 기반으로 사전에 가짜뉴스라고 판별된 뉴스와 유사한 내용을 트위터에서 수집하고 그것이 진짜인지 가짜인지 라벨링을 진행하는 것으로 데이터셋을 구축했다. 아울러 코로나19 관련 한국어 가짜뉴스 탐지 연구에 관심을 갖고 있는 다른 연구자들도 본 데이터를 이용할 수 있도록 구축된 데이터셋을 Github에 공개하였다.

데이터셋을 구축한 이후, 본 연구에서는 한국어 형태소 분석기인 Okt를 활용해 각 트윗별로 각 품사가 얼마나 자주 출현했는지를 계량화하여 해당 데이터셋으로부터 언어학적 특징변수를 추출하고, 이를 가짜뉴스 탐지모델에 활용하고자 하였다. 그런 다음 Doc2Vec으로 임베딩된 텍스트 정보만 활용했을 경우와 비교해 품사별 출현 빈도 정보를 추가로 활용했을 때 예측성능이 제고되는지 확인해 보고자 하였다. 실험 결과, CatBoost를 포함한

모든 기계학습 기법에서 F1 점수가 상승한 것을 확인할 수 있었다. 가장 큰 F1 점수의 상승폭을 보인 기계학습 기법은 Logistic Regression으로 나타났으며, 비교모델에 비해 제안모델이 약 7.5% 정도의 상승폭을 보이는 것을 확인할 수 있었다. 특히 큰 변동이 없었던 Precision에 비해 Recall값이 크게 향상되어, 실제 가짜뉴스를 제안 모델이 가짜뉴스라고 판별하는 능력이 괄목할 수준으로 향상되었음을 확인할 수 있었다.

현윤진과 김남규(2018)는 언어적 특성이 콘텐츠에 크게 의존하는 트위터에서 가짜뉴스를 유포하는 사람이 실제 뉴스와 유사한 언어 형식으로 작성할 경우 내용 기반의 가짜뉴스 탐지 기법의 성능이 떨어질 수 있다고 문제를 제기한 바 있다. 본 연구에서는 품사의 빈도 정보를 활용한 것만으로도 성능의 개선이 이루어질 수 있음을 확인함으로써 현윤진과 김남규(2018)의 연구에서 제기한 문제를 해결할 수 있는 단초를 제시했다는 점에서 학술적 의의를 갖는다. 또한 영어를 모국어로 사용하는 영미권 국가에서는 다수의 가짜뉴스 공개 데이터셋을 활용하여 다양한 연구들이 활발하게 진행 중이나 국내에서는 한국어에 특화된 가짜뉴스 탐지 연구가 부족하고 데이터도 모자란 상황인데, 본 연구에서는 직접 수집한 데이터셋을 Github을 통해 공유함으로써 국내 가짜뉴스 탐지 연구 활성화에 기여하였다는 점에서도 의의를 찾을 수 있다.

이러한 학술적 의의를 갖고 있지만, 사용한 데이터셋의 크기가 너무 작다는 점은 본 연구의 피할 수 없는 중대한 한계점이라 할 수 있다. 소수의 인원으로 구성된 연구진이 독립적으로 자료 조사와 수집을 통해 가짜뉴스 실험 데이터셋을 구축해야 했기에 충분한 분량의 한국어 데이터를 확보하는 데는 현실적으로 어려움이 있었다. 하지만, 크기가 작은 데이터셋은 모델의 성능을 평가했을 때 예측 결과가 정확하지 않을 수 있고 과적합이 발생할 확률이 높기 때문에(신성윤 등, 2020), 향후 지속적인 업데이트를 통해 데이터셋의 크기를 키울 필요가 있다.

또한 김수연 등(2022)에 의하면 본 실험에 사용한 Okt 형태소 분석기는 용언과 어미를 명확하게 구분하지 못해 다른 형태소 분석기에 비해 POS Tagging의 정확도가 떨어질 가능성이 있다. 따라서 향후 연구에서는 품사 정보를 정확히 추출할 수 있는 형태소 분석기를 물색하고 이에 대한 적용을 검토할 필요가 있다.

끝으로 본 연구에서는 가짜뉴스의 내용(contents)만 활용하고, 배경정보(context)의 활용을 배제하였다. 그러나 최근에는 내용 기반의 가짜뉴스 탐지보다 배경정보를 통한 소셜 컨텍스트 기반의 가짜뉴스 탐지 기법의 연구가 더 주목받고 있다. 이러한 최근의 연구동향을 반영하여, 향후 연구에서는 배경정보를 추가로 활용함으로써 콘텐츠에 의존하는 내용 기반 탐지 기법의 개선을 도모할 필요가 있다. 아울러 최근에는 동영상 형태로 된 유튜브 가짜뉴스가 사회적으로 큰 영향을 미치고 있는데, 내용 기반 탐지를 적용하는데 있어서도 텍스트뿐 아니라 이미지, 동영상 등 멀티미디어 내용을 종합적으로 고려하는 멀티모달(multi-modal) 접근 도입을 적극 검토해 보아야 할 것으로 사료된다.

참고문헌(References)

[국내 문헌]

국립재난안전연구원. (2020). *Future Safety Issue 제2의 팬데믹 인포데믹으로 인한 혼돈의 시대*, from https://www.ndmi.go.kr/home/sub.do?menukey=6031&mode=view&no=1316137 (2020/10/12)

길호현. (2018). 텍스트마이닝을 위한 한국어 불용어 목록 연구. *우리말글*, *78*, 1-25.

김수연, 안석호, 김동현, 이의종, 서영덕. (2022, June). 형태소 분석기의 품사별 정확성 분석. *In Proceedings of KIIT Conference* (pp. 378-381).

민희. (2022). 가짜 뉴스 확산, 그 이후: 보수와 진보의 가짜 뉴스 노출과 제도 신뢰의 편향. *정치정보연구*, *25*(3), 151-180.

박수선. (2023, January 30). 예산 대폭 깎인 팩트체크넷, 결국 문 닫는다, PD저널, Available at http://www.pdjournal.com/news/articleView.html?idxno=74609(Accessed 2023. 4. 19)

손승혜, 이귀옥, 홍주현, 최지향, 정은정. (2018). 트위터는 어떻게 가짜 뉴스를 유통시키는가?: <교통법규 개정설>과 <9월 전쟁설>의 트위터 유통 패턴과 유력자, 빈출단어 분석. *사이버커뮤니케이션학보*, *35*(4), 203-251.

송찬우, 안현철. (2022). Tag2vec 기반의 지능형 불법 도박 사이트 탐지 모형 개발. *지능정보연구*, *28*(4), 211-227.

신성윤, 신광성, 이현창. (2020). 적은 데이터 세트를 기반으로 한 동물 이미지의 향상된 딥 러닝. *한국컴퓨터정보학회 학술발표논문집*, *28*(1), 247-248.

심재승, 이재준, 정이태, 안현철. (2020). 워드 임베딩을 활용한 한국어 가짜 뉴스 탐지 모델에 관한 연구. *한국컴퓨터정보학회 학술발표논문집*, *28*(2), 199-202.

심홍진, 황유선. (2010). 마이크로블로깅 (micro-blogging) 이용동기에 관한 연구: 트위터 (twitter) 를 중심으로. *한국방송학보*, *24*(2), 192-234.

안형준. (2020). SNS 의 이벤트와 텍스트의 언어 심리학적 특성 간의 관계. *한국정보기술학회논문지*, *18*(5), 91-100.

원혜진, 이현영, 강승식. (2020). 대규모 텍스트 분석을 위한 한국어 형태소 분석기의 실행 성능 비교. *한국정보과학회 학술발표논문집*, 401-403.

이장근, 김해연, 장적, 김용환. (2022). 가짜 뉴스 영향력 인식의 효과에 관한 연구: 매체별 가짜 뉴스 제 3 자 인식이 가짜 뉴스 규제 및 미디어 교육 필요성에 미치는 영향. *한국콘텐츠학회논문지*, *22*(12), 316-326.

이종구. (2020, March 22). *방역한다며 메탄올 뿌린 뒤 중독 증세...'정보 전염병' 피해*, YTN, Available at https://www.ytn.co.kr/_ln/0103_202003222000523006_018 (Accessed 2023. 4. 19)

이지민, 이지선, 우지영. (2022). 코로나 19 가짜 뉴스와 진짜 뉴스 판별 시스템. *한국컴퓨터정보학회 학술발표논문집, 30(1),* 411-412.

정세민, 이세영, 안유나, 김보경. (2021, November). 품사에 따른 영화 리뷰 감성분석 연구. In Proceedings of KIIT Conference, 651-654.

정예림, 김지희, 유형선. (2020). Word2Vec 을 활용한 제품군별 시장규모 추정 방법에 관한 연구. *지능정보연구*, *26*(1), 1-21.

정이태, 안현철. (2022). 그래프 임베딩을 활용한 코로나 19 가짜 뉴스 탐지 연구-사회적 참여 네트워크의 이용 여부에 따른 탐지 성능 비교. *지능정보연구*, *28*(1), 197-216.

한국언론진흥재단_뉴스빅데이터_메타데이터_가짜뉴스, Retrieved from https:// www.data.go.kr/data/15086437/fileData.do (Accessed 2023. 4. 19)

한국언론진흥재단_뉴스빅데이터_메타데이터_코로나, Retrieved from https:// www.data.go.kr/data/15069309/fileData.do (Accessed 2023. 4. 19)

한소은, 강윤석, 고윤용, 안지원, 김유심, 오성수, 박희진, 김상욱. (2022). CoAID+: 소셜 컨텍스트 기반 가짜 뉴스 탐지를 위한 COVID-19 뉴스 파급 데이터. *정보처리학회논문지. 소프트웨어 및 데이터 공학*, *11*(4), 149-156.

한지원, 김영욱. (2023). 댓글의 방향과 강도가 코로나 19 관련 가짜 뉴스 수용에 미치는 영향: 체계적 정보처리의 매개효과 및 동조성향의 조절효과 중심 분석. *한국언론학보*, 67(1), 230-271.

현윤진, 김남규. (2018). 뉴스와 소셜 데이터를 활용한 텍스트 기반 가짜 뉴스 탐지 방법론. *한국전자거래학회지*, *23*(4), 19-39.

Open-Korean-Text(OKT), Retrieved from https://github.com/open-korean-text/open-korean-text (Accessed 2023. 4. 19)

[국외 문헌]

Bondielli, A., & Marcelloni, F. (2019). A survey on fake news and rumour detection techniques. *Information Sciences*, 497, 38-55.

Bovet, A., & Makse, H. A. (2019). Influence of fake news in Twitter during the 2016 US presidential election. *Nature Communications*, 10(1), 7.

Cui, L., & Lee, D. (2020). Coaid: Covid-19 healthcare misinformation dataset. arXiv preprint arXiv:2006.00885.

Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., ... & Zittrain, J. L. (2018). The science of fake news. *Science*, 359(6380), 1094-1096.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013, December). Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems-Volume 2* (pp. 3111-3119).

Ngada, O., & Haskins, B. (2020, December). Fake news detection using content-based features and machine learning. In *2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)* (pp. 1-6). IEEE.

Patwa, P., Sharma, S., Pykl, S., Guptha, V., Kumari, G., Akhtar, M. S., ... & Chakraborty, T. (2021). Fighting an infodemic: Covid-19 fake news dataset. In *Combating Online Hostile Posts in Regional Languages during Emergency Situation: First International Workshop, CONSTRAINT 2021*, Collocated with AAAI 2021, Virtual Event, February 8, 2021, Revised Selected Papers 1 (pp. 21-29). Springer International Publishing.

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, 19(1), 22-36.

Tandoc Jr, E. C., Lim, Z. W., & Ling, R. (2018). Defining “fake news” A typology of scholarly definitions. *Digital Journalism*, 6(2), 137-153.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146-1151.

Abstract

COVID-19-related Korean Fake News Detection Using Occurrence Frequencies of Parts of Speech

Jihyeok Kim* · Hyunchul Ahn**

The COVID-19 pandemic, which began in December 2019 and continues to this day, has left the public needing information to help them cope with the pandemic. However, COVID-19-related fake news on social media seriously threatens the public's health. In particular, if fake news related to COVID-19 is massively spread with similar content, the time required for verification to determine whether it is genuine or fake will be prolonged, posing a severe threat to our society. In response, academics have been actively researching intelligent models that can quickly detect COVID-19-related fake news. Still, the data used in most of the existing studies are in English, and studies on Korean fake news detection are scarce. In this study, we collect data on COVID-19-related fake news written in Korean that is spread on social media and propose an intelligent fake news detection model using it. The proposed model utilizes the frequency information of parts of speech, one of the linguistic characteristics, to improve the prediction performance of the fake news detection model based on Doc2Vec, a document embedding technique mainly used in prior studies. The empirical analysis shows that the proposed model can more accurately identify Korean COVID-19-related fake news by increasing the recall and F1 score compared to the comparison model.

Key Words : COVID-19-related Fake News, Korean Fake News, Social Media, Doc2Vec, POS Tagging

Received : May 15, 2023 Revised : June 12, 2023 Accepted : June 19, 2023
Corresponding Author : Hyunchul Ahn

* Graduate School of Business IT, Kookmin University
** Corresponding author: Hyunchul Ahn
Graduate School of Business IT, Kookmin University
77, Jeongneung-ro, Seongbuk-gu, Seoul 02707, Korea
Tel: +82-2-910-4577, Fax: +82-2-910-4017, E-mail: hcahn@kookmin.ac.kr

저 자 소 개

김 지 혁

현재 국민대학교 비즈니스IT전문대학원 석사과정에 재학 중이며, 선문대학교 IT경영학 전공으로 학사 학위를 취득하였다. 주요 관심분야는 Big Data Analytics, Business Analytics 등이다.

안 현 철

현재 국민대학교 비즈니스IT전문대학원 교수로 재직 중이다. KAIST에서 산업경영학사를 취득하고, KAIST 테크노경영대학원에서 경영정보시스템을 전공하여 공학석사와 박사를 취득하였다. 주요 관심 분야는 금융 및 고객관계관리 분야의 인공지능 응용, 지능형 의사결정지원시스템, 정보시스템 수용과 관련한 행동 모형 등이다.