Information Systems Review
Vol. 23, No. 4 November 2021 http://dx.doi.org/10.14329/isr.2021.23.4.127

텍스트 마이닝과 딥러닝 알고리즘을 이용한 가짜 뉴스 탐지 모델 개발

Development of a Fake News Detection Model Using Text Mining and Deep Learning Algorithms

임 동 훈 (Dong-Hoon Lim) ㈜데이터월드 차장
김 건 우 (Gunwoo Kim) 한밭대학교 융합경영학과 교수
최 근 호 (Keunho Choi) 한밭대학교 융합경영학과 조교수, 교신저자

요 약

가짜 뉴스는 정보화 시대라는 현대사회의 특성에 의해 진위 여부의 검증과는 상관없이 빠른 속도로 확대, 재생산되어 퍼진다. 전체 뉴스의 1%를 가짜라고 가정했을 경우 우리사회에 미치는 경제적 비용이 30조 원에 달한다고 하니 가짜 뉴스는 사회적, 경제적으로 매우 중요한 문제라고 할 수 있다. 이에 본 연구는 뉴스의 진위 여부를 신속하고 정확하게 확인하고자 자동화된 가짜 뉴스 탐지 모델을 개발하는데 목적을 두고 있다. 이를 위해 본 연구에서는 크롤링(crawling)을 통해 진위 여부가 밝혀진 뉴스 기사를 수집하였고, 워드 임베딩(Word2Vec, Fasttext)과 딥러닝 기법(LSTM, BiLSTM)을 이용하여 가짜 뉴스 예측 모델을 개발하였다. 실험 결과, Word2Vec과 BiLSTM의 조합이 가장 높은 84%의 정확도를 보였다.

키워드 : 가짜 뉴스, 한국어 뉴스, 자연어 처리, 딥러닝, 텍스트 마이닝

Ⅰ. 서 론

기근, 질병, 전쟁, 사회 혼란 등 사람들의 불안이 커지는 '위기의 시기'에는 가짜 뉴스가 기승을 부린다. 오늘날 코로나19와 같은 세계적 감염병의 시기에 예외 없이 가짜 뉴스가 만연하다는 점은 더 이상 놀라운 사실이 아니다. 가짜 뉴스의 생산과 유통은 전혀 새로운 것이 아니지만 정보통신 기술의 발달과 사회관계망 서비스 등의 성장에 따라 과거보다 가짜 뉴스 확산의 속도는 빨라지고 범위는 넓어지고 있다. WHO는 코로나19와 같은 세계적 감염병을 경고하면서 동시에 인포데믹에 대한 경고도 하고 있다. 미국에서는 2020년 3월 11일~4월 20일 동안 306건의 코로나19관련 가짜 뉴스(비방 37%, 의학정보 30%, 방역정책 8%)가 조사되었고(Poynter 연구소 산하 코로나바이러스 팩트 체크 연합), 국내에서도 2020년 1월 28일~4월 24일 동안 96건의 코로나19 관련 가짜 뉴스(의학정보 52%, 비방 33%, 방역정책 8%)가 조사되었다(성욱제, 정은진, 2020).

대선과 같은 국가의 큰 정치적 이벤트가 있을 경우에도 가짜 뉴스는 기승을 부린다. 2017년 프

랑스 대선에서 당시 대선후보였던 에마뉘엘 마크롱에 대해 '사우디아라비아 정부가 마크롱의 캠페인에 돈을 대주고 있다.'는 기사가 실렸는데, 해당 기사는 벨기에의 대표적인 언론 르수아(Le Soir)를 모방한 가짜 웹사이트에 게재된 가짜 뉴스였다(진민정, 2017). 그 보다 1년 앞선 2016년 미국 대선에서는 소셜미디어를 활용한 가짜 뉴스가 논쟁의 중심이었다. 선거 여론에 영향을 미치는 가짜 뉴스들이 페이스북(Facebook) 등을 통해 급속하게 확산되면서 그 배경과 함께 실제 여론의 왜곡 현상으로 이어졌는지가 논란이 되었다. 미국의 인터넷 뉴스매체인 버즈피드(BuzzFeed)에 따르면 대선 전 3개월간 가장 인기가 있었던 가짜 뉴스 20개의 페이스북 내 공유, 반응, 댓글 수는 총 871만 1천 건에 달했다. 이는 CNN, 뉴욕타임즈 등 주요 전통미디어의 가장 반응이 높았던 대선 기사 20개의 반응(736만 건)을 넘어선 수치였다(김유향, 2016).

해당 가짜 뉴스를 게시한 언론사 중 Ending the Fed는 트럼프 지지 세력이었으며, 그 외 다른 언론사들은 클릭 광고로 인한 수익 창출이 목적인 것으로 밝혀졌다. 이처럼 정치적 목적이 아닌 단순한 관심을 끌거나 광고 수입을 목적으로 한 가짜 뉴스도 급증하고 있다. 이러한 가짜 뉴스로 인해 개인과 기업, 국가는 많은 사회적 비용을 치르고 있다. 전체 뉴스의 1%가 가짜 뉴스라고 가정하였을 때 이러한 가짜 뉴스로 우리사회가 치러야 할 경제적 비용은 약 30조 900억 원에 달하는 것으로 추정된다(주원 등, 2017).

현재까지 가짜 뉴스에 대한 정의는 학술적으로 합의되지 않고 있으나, Allcot and Gentzkow(2017)는 "독자들이 오해할 고의적이고 검증 가능한 뉴스 기사", 가디언지는 "트래픽과 이윤을 극대화하기 위해 독자들을 속이기 위해 설계되고 만들어진 기사"로 정의하였다. 국내 언론에서는 "전체 또는 일정 부분이 사실이 아닌 정보에 근거해 만들어진 기사나 뉴스 형태"라고 정의한 바 있다(뉴스케어, 2017). 그 외에도 "상업적 또는 정치적 목적을 가진 기만적인 정보"(황용석, 2017) 등 가짜 뉴스는 처음에는 뉴스 기사의 형식을 빌린 것을 칭하였으나, 점점 실제로 모든 허위사실을 가리키는 용어로 바뀌어 가고 있다(양정애, 2019).

가짜 뉴스가 사회적, 정치적 문제로 많은 관심을 받게 되면서 다양한 곳에서 가짜 뉴스에 대한 해법을 제시하고 있다. 이를 정리해보면, 가짜 뉴스를 탐지하는 방법은 비기술적 방법과 기술적 방법, 그리고 두 가지 이상의 기법을 활용하는 하이브리드 기법으로 분류할 수 있다. 현재는 비기술적 접근이 주류를 이루고 있지만 가짜 뉴스를 소수의 전문가가 모두 확인하기는 현실적으로 어려운 한계가 있다. 최근에는 기술적 접근 기법을 통해 가짜 뉴스를 자동으로 탐지하는 시스템에 대한 많은 연구가 이루어지고 있다(윤영석 등, 2017).

본 연구는 가짜로 판명된 뉴스와 해당 뉴스의 진짜(검증)뉴스를 학습데이터로 사용하여 인공지능 기반의 한국어 뉴스 진위 여부 분류 모델을 개발함으로써 가짜뉴스 판별에 도움을 주는 것을 목적으로 한다.

본 연구에서 사용한 분석 대상의 범위는 서울대학교 언론정보연구소의 팩트체크(SNU Factcheck) 사이트 https://factcheck.snu.ac.kr에 게시된 한국어 뉴스 중 '전혀 사실이 아님', '대체로 사실이 아님' 배지(Badge)를 받은 뉴스와 해당 뉴스의 검증 뉴스이다. 본 연구는 분류를 위한 기법으로 딥러닝 기법 중 하나인 Long Short-Term Memory(LSTM)과 Bidirectional LSTM(BiLSTM)을 사용하였으며, 모델의 성능평가를 위한 지표로 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), 그리고 F1 Score를 사용하였다. 또한 정확도 향상을 위해, 선행연구에서 많이 활용되지 않았던 뉴스의 제목, 뉴스 기사의 주제, 뉴스 기사의 상세 주제, 뉴스의 중심이 되는 대상, 뉴스의 중심 내용을 주장 또는 검증한 주체, 그리고 뉴스를 주장 또는 검증한 매체 등의 다양한 메타 정보를 독립변수에 추가하여 모델을 개발하였다.

본 논문의 구성은 다음과 같다. 제Ⅱ장에서는

자연어 처리, 자동화 기반 가짜 뉴스 탐지, 딥러닝 기법 등 관련 선행 연구에 대해 소개하고, 제Ⅲ장에서는 연구 프레임워크와 분석 데이터 등 연구방법을 설명한다. 제Ⅳ장에서는 연구 결과에 대해 설명하고, 마지막으로 제Ⅴ장에서는 본 연구의 결론과 시사점, 그리고 한계점과 더불어 향후 연구 방향에 대해 제시한다.

Ⅱ. 관련연구

2.1 워드 임베딩(Word Embedding)

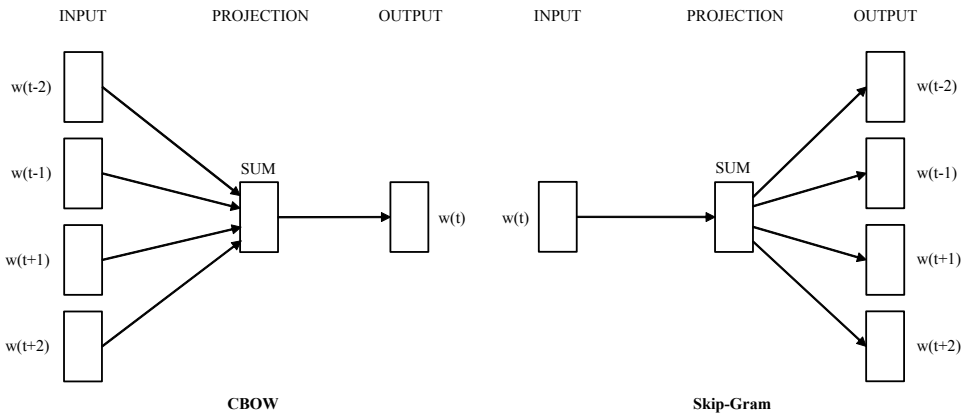
기계의 학습은 여러 연산을 통해 이루어진다. 자연어로 이루어진 텍스트를 그대로 기계에게 전달하여 연산을 수행시킬 수는 없다. 텍스트를 숫자로 변환하여야 알고리즘에 의해 계산을 할 수 있는데, 텍스트를 숫자로 바꾸는 방법 중 하나로 단어를 벡터로 변환하는 방법이 있다. '야구', '축구', '농구'라는 단어들을 벡터로 변환하려면 각 단어에 해당하는 요소만 1로 하고, 나머지를 0으로 채울 수 있다. 즉, '야구'는 [1, 0, 0], '축구'는 [0, 1, 0], '농구'는 [0, 0, 1]과 같이 표현이 가능한데, 이러한 방식을 one-hot encoding이라고 한다. One-hot encoding은 단어를 벡터로 변환하기 편리하지만 활용하는데 있어 몇 가지 국소표현(Local Presentation)의 문제가 발생할 수 있다. 만약 n개의 단어가 있을 경우 n차원의 벡터로 표현을 해야 하는데 단어의 수가 많아지면 벡터의 차원이 너무 커지게 된다(Long Vectors). 또한 단어 간의 유사도를 알 수 없기에 단어의 의미를 알 수가 없다(Sparse Representation). 단어 임베딩 모델(Word Embedding Model)은 단어를 벡터로 변환할 때 벡터에 단어의 의미를 담을 수 있도록 변환하는 모델이며, Word2vec, Fasttext는 그 중 대표적인 모델이라 할 수 있다(이기창, 2019).

Word2vec은 Google에서 2013년도에 제안한 방법론으로 유사한 문맥을 가진 단어들은 의미도 유사하게 가진다는 언어학의 분포가설(Distributional Hypothesis)에서 출발하였다. 단어의 의미를 벡터 공간에 임베딩 하여 표현하는 방식으로, 한 단어의 주변 단어들을 연관된 단어로 가정하여 벡터 공간에서의 거리를 점차 줄임으로서 그 단어의 의미를 파악하는 방법이다. 학습모델은 <그림 1>에서 보는 바와 같이 CBOW(Continuous Bag of Words)와 Skip-gram이 있다. CBOW는 맥락(Context)을 통해 타겟 단어(Target word)를 예측하는데, 맥락이란 주변 단어(앞, 뒤의 단어)를 통해 알 수 있는 정보를 뜻한다. 이 주변 단어의 범위를 window라고 하며, 몇 개의 단어를 통해 예측할 것인가를 지정하는 것을 window size라고 한다. Skip-gram은 CBOW와 반대로 타겟 단어(중심단어)를 통해 주변 단어를 예측하는 방식이다. 예를 들어, '나는 따듯한 밥을 먹었다.'라는 문장에서 CBOW는 "나는", "따듯한", "먹는다"라는 주변 단어들을 통하여 "밥을"이라는 단어를 예측하는 방식이라면, Skip-gram은 "밥을"이라는 중심단어를 통해 주변 단어들을 예측하는 방식이다.

$$\frac{1}{V}\sum_{c=1}^{V}\sum_{\substack{-m\le j<m\\ j\neq 0}} \log p(w_{c+j}|w_c) \qquad (1)$$

$$p(w_o \mid w_c) = \frac{\exp(v_{w_o^T} u_{w_o})}{\sum_{w=1}^{V} \exp(v_{w_o^T} u_{w_o})} \qquad (2)$$

Skip-Gram은 식 (1)과 같이 모수 W, W'를 추정하는 것을 목적으로 하며, 타겟 단어를 중심으로 주변 단어들을 예측하는 확률인 $p(W_{c+j}|W_c)$는 Softmax함수를 통해 식 (2)와 같이 표현할 수 있다. W_o는 주변단어, W_c는 중심단어를 뜻한다. 일반적으로 Skip-gram 모델이 CBOW 모델보다 더 좋은 성능을 보이는 것으로 알려져 있는데(채상희, 2019) 주변단어로부터 중심단어를 예측하는 CBOW의 경우는 중심단어가 한 번의 업데이트 기회를 갖는 반면, 중심단어로부터 주변단어를 예측하는 Skip-gram의 경우 window size만큼 중심단어의 업데이트가 일어나게 되어 학습량이 CBOW보다 많기 때문이다.



〈그림 1〉 CBOW, Skip-Gram 네트워크 구조(Yuan *et al.*, 2020)

Fasttext는 facebook에서 2016년도에 제안한 방법론으로, 기존 임베딩 모델의 몇몇 한계점을 개선하여 나왔다. 기존 단어 임베딩 모델에서는 단어를 개별적으로 임베딩하여 단어의 형태학적(Morphological) 특성을 반영하지 못하였고, 출현빈도가 적은 희소단어(Rare word)에 대한 임베딩이 어려웠다. 또한, 학습용 말뭉치(Corpus)에 존재하지 않는 새로운 단어(OOV, Out Of Vocabulary)는 처리하지 못하는 한계점도 있었다. Fasttext는 이러한 한계점을 보완하여 나왔는데, 원래 단어를 부분단어(Subword)의 벡터로 표현하는 점 외에는 Word2vec의 Skip-gram과 비슷하다(조현수, 이상구, 2017).

$$p(w_o \mid w_c) = \frac{\exp(s(w_c, w_o))}{\sum_{j=1}^{V} \exp(s(w_c, w_j))} \quad (3)$$

기존 Word2vec이 'where'이라는 단어를 하나의 어휘로 보고 임의의 벡터를 할당하고 학습하였다면, Fasttext는 <wh, whe, her, ere, re>로 5개의 n-gram(n=3)으로 분리 후 임베딩 벡터를 할당하여 평균 벡터로 계산한다(채상희, 2019). 위 알고리즘을 요약하면 <표 1>과 같다.

2.2 자동화 기반 가짜뉴스 탐지 관련 연구

앞서 서론에서 서술하였듯이 자동화된 가짜뉴스 탐지를 위해, 인공지능 기반이나, 시맨틱 기반, 그리고 이상 확산패턴 탐지와 같은 기술적 접근 기법이 많이 활용되고 있다. 인공지능 기반은 가짜뉴스로 판정된 이전 뉴스들에서 빈번하게 사용된 단어와 표현을 분석하여 도출된 정보를 기계학습을 통해 학습하여 모델을 생성하고, 해당 모델을 기반으로 새로운 뉴스가 진짜인지 가짜인지에 대한 확률을 예측하는 기법이다(윤영석 등, 2017). 신속한 분석을 통해 가짜여부를 빠르게 판단할 수 있기 때문에 자동화된 가짜뉴스 탐지 시스템에서 최근 많이 활용하고 있다.

〈표 1〉 임베딩 방식 요약

임베딩 모델		설명
Word2vec	Skip-gram	중심단어로부터 window size 내 주변단어를 예측
	CBOW	Window size 내 주변단어로부터 중심단어를 예측
Fasttext		원래 단어를 부분단어의 벡터로 표현

윤태욱(2018)의 연구에서는 뉴스제목과 관련 메타 정보를 활용하여 가짜뉴스를 탐지하고자 하였다. SNU Factcheck에서 2017년 3월~2017년 9월까지 150건(진짜 50건, 중립 50건, 가짜 50건)의 뉴스에 대한 제목과 출처 등의 메타 정보를 수집하였고, 7:3 비율로 학습 데이터와 검증 데이터를 구성하여 연구를 진행하였다. 뉴스 제목을 꼬꼬마 형태소분석기와 Term Frequency-Inverse Document Frequency(TF-IDF)를 사용하여 토픽 모델링을 수행하였고, 여러 메타 정보 중 최종적으로 언론사를 독립변수로 추가하였다. Support Vector Machine(SVM), Artificial Neural Network(ANN), Case-Based Reasoning(CBR), Mean Decrease in Accuracy(MDA)를 활용하여 다분류 예측 모델을 만들어 5-fold 교차검증으로 실험을 진행하였으며, 뉴스제목 + 언론사 + SVM 조합이 검증 평균 정확도 55.33%로 가장 높은 정확도를 보였다.

현윤진, 김남규(2018)의 연구에서는 뉴스와 소셜 데이터를 활용하여 가짜뉴스를 탐지하고자 하였다. 2017년 1월~2018년 6월까지 서울대학교 언론정보연구소 팩트체크, 뉴스톱, 네이버뉴스에서 총 134건(진짜 78건, 가짜 56건)을 수집하였으며, 2016년 1월~2018년 7월 사이 16,384건의 트위터 내용을 수집하였다. 그 중 77건을 학습, 57건을 검증데이터로 분할하였고, 트위터는 10,371건을 학습, 6,013건을 검증용으로 사용하였다. 신경망 알고리즘으로 분류를 수행하였고, 정확도는 뉴스만 활용하였을 경우 52.63%, 트위터만 활용하였을 때 56.14%, 두 개를 결합하였을 때는 80.7.%의 정확도를 보였다.

Nikam and Dalvi(2020)은 Kaggle에서 수집한 Twitter 3,983건(진짜 2,118건, 가짜 1,865건)을 학습, 검증용으로 각각 7:3으로 분할하여 TF-IDF로 토픽 모델링을 수행 후 Naïve Bayes, Passive aggressive Classifier를 이용하여 분류 모델을 개발하였다. 그 결과 Naïve Bayes는 73%, Passive aggressive는 78%의 정확도를 보였다. 위의 기존 연구들을 정리하면 <표 2>와 같다.

또한 최근에는 가짜뉴스 탐지를 위한 다양한 챌린지가 개최되고 있다. 딘 포멀로(Dean Pomerleau)와 델립 라오(Delip Rao)에 의해 추진된 'Fake News Challenge(http://www.fakenewschallenge.org/)'는 전 세계 학회와 업계 100여 명이 넘는 자원 봉사자와 71개 팀이 참가하였다. 이 챌린지의 목표는 기계학습과 자연어 처리, 인공지능 같은 기술을 활용해 뉴스 기사에 숨겨진 조작, 오보를 식별할 수 있는 가능성을 모색하는 것으로 2016년 12월부터 시작하여 2017년 6월까지 진행되었으며 최종 상위 3개팀에서 딥러닝, 멀티스레딩, TF-IDF, 의사결정나무 등의 기법을 이용하였다. 국내에서는 2017년 인공지능 R&D 챌린지가 열려 대학 14개 팀, 기업 28개 팀, 연구소 3개 팀, 개인 26개 팀이 참가하여 가짜뉴스 찾기에 도전하였다. 해당 챌린지는 제목과 본문 내용의 적합성을 판별하는 임무1과 본문 중 맥락에 관계없는 내용이 있는지를 찾는 임무2로 나뉘어 진행되었다. 평가지표는 AUCROC를 기준

〈표 2〉 기존 자동화 기반 가짜 뉴스 분류 연구

연구자	데이터 출처	데이터 구성	방법론	결과(정확도)
윤태욱(2018)	SNU, FactCheck	뉴스: 진짜, 중립, 가짜 각 50건	토픽모델링 + SVM	55.33%
현윤진, 김남규(2018)	SNU, FactCheck, 트위터	뉴스: 진짜 78건, 가짜 56건 트위터: 10,371건	토픽모델링 + Neural Network	뉴스: 52.63% 트위터: 56.14% 뉴스+트위터: 80.7%
Nikam and Dalvi(2020)	트위터	트위터: 진짜 2,118건, 가짜 1,865건	토픽모델링 + Naïve Bayes Classifier 토픽모델링 + Passive Aggressive Classifier	NB: 73% PA: 78%

으로 하였고, 실제 한국어뉴스 1만여 건이 활용되었다. 최종 상위 3개 팀에서는 워드 임베딩과 심층학습을 접목하였는데, 자체 개발한 검색엔진과 형태소·구문 분석 등 다양한 자연어처리 기술을 활용하였으며, 심층학습 기법으로 가짜뉴스를 판별했다(좌희정 등, 2019).

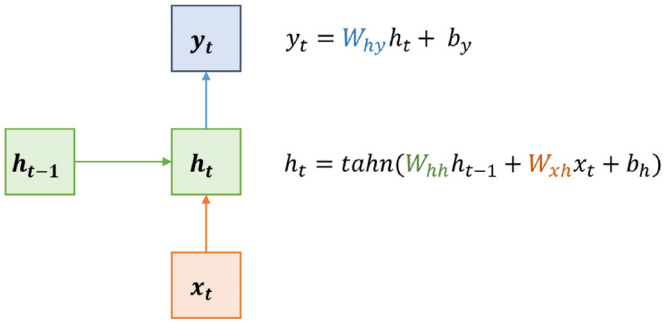
2.3 딥러닝 기법

2.3.1 Long Short-Term Memory(LSTM)

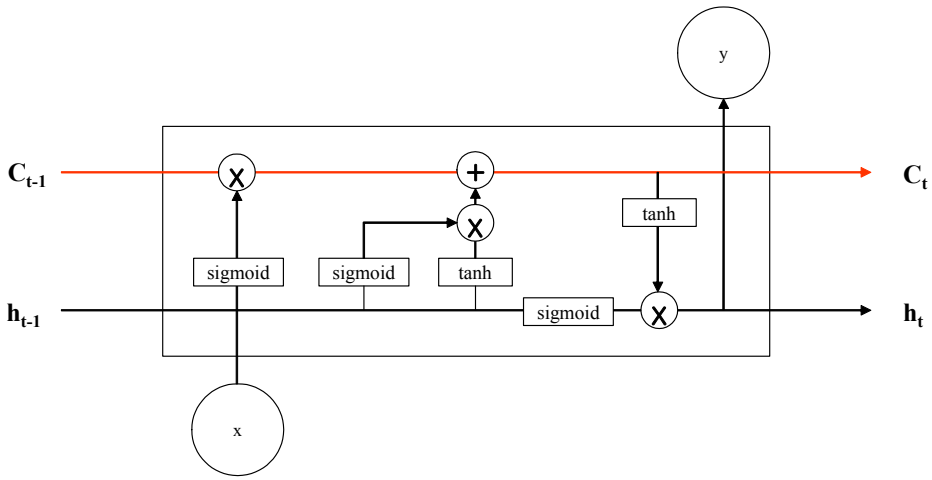
"traffic ticket fines", "traffic is fine" 두 문장은 중간에 ticket이라는 단어를 중심으로 앞뒤로 같은 단어를 사용하는 유사한 문장이다. 하지만 저 문장들에 대한 감성을 분류해본다면 첫 번째 문장은 "교통 벌금 티켓"이라는 뜻이므로 부정적인 감정에 가깝고, 두 번째 문장은 "교통이 괜찮다"라는 의미이므로 긍정에 가깝다. 같은 traffic으로 시작하였지만 뒤에 단어들이 ticket fines가 붙게 되면 부정적, is fine이 붙게 되면 긍정적인 감성을 보이는데 이렇게 단어의 연결관계에 의해 감성을 판단할 수 있다. <그림 2>에서 보는 바와 같이, 입력되어지는 단어들의 앞뒤 연결 관계의 가중치를 통해서 타겟을 예측, 학습할 수 있도록 한 인공 신경망인 Recurrent Neural Network(RNN)은 데이터가 순차적으로 입력될 때 이전 단계에서 전달받은 가중치를 통해 현재 단계를 수행하고, 다음 단계로 전달하는 방식이다(Williams *et al.*, 1986).

본 연구에서는 뉴스 기사의 진위여부를 분류하고자 하였는데, 이처럼 진짜/가짜를 분류하는 것은 긍정/부정을 분류하는 이진분류와 동일하다. 따라서, RNN을 분석에 활용할 수 있다. 하지만 뉴스 기사는 한, 두 문장이 아닌 여러 문장으로 이루어진 긴 글로서, 문장표현의 순서상 위치가 멀어질수록 RNN은 문맥정보를 연결하기 어려워진다. 학습이란 최적의 가중치(w)를 찾는 과정인데, 이때 사용되는 방법이 경사하강법(Gradient Descent)이다. 새로운 가중치는 기존의 가중치에서 학습률 곱하기 Error를 가중치로 미분한 값으로 구하게 되는데 ($w = w - \alpha \times \frac{\partial E}{\partial w}$) 이때 sequence가 길게 되면 Gradient Vanishing 기울기가 0으로 수렴하여 가중치에 대한 업데이트가 이루어지지 않는 현상, Gradient Exploding 가중치의 업데이트가 크게 이루어져 학습이 안정적이지 않게 되는 현상 이슈가 발생하게 된다. 이러한 문제를 장기 의존성 문제라고 하는데 이것을 개선한 모델이 LSTM(Long Short-Term Memory)이다(Hochreiter and Schmidhuber, 1997).

LSTM은 <그림 3>에서 보는 바와 같이, 먼저 이전 단계에서 전달된 정보 중 얼마만큼을 현재에 반영할 것인지를 결정하고, 현재 단계에서 입력된 값을 계산하여 만들어진 Memory cell(C_t)을 다음 단계로 전달하도록 되어있다. 따라서 RNN처럼 입력 차이가 큰 데이터의 연관성이 사라지는 장기 의존성 문제가 발생하지 않는다(유은조 등, 2018).



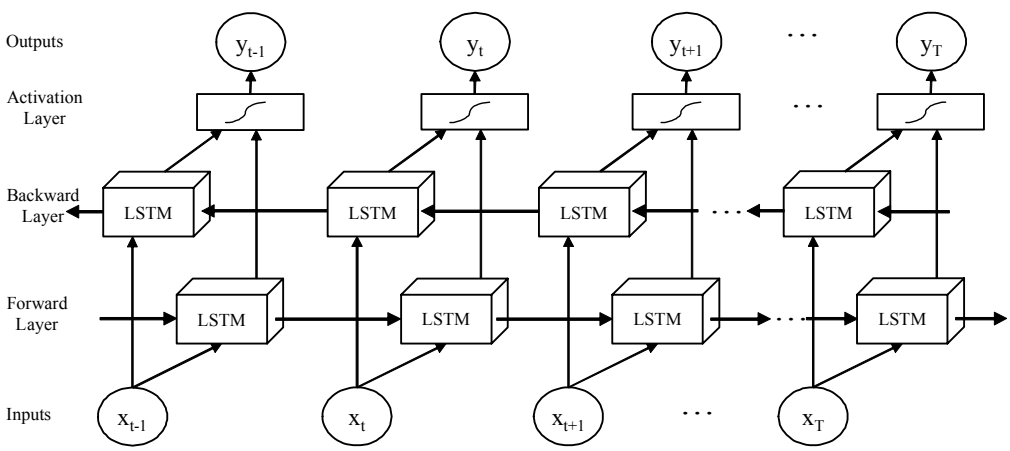
〈그림 2〉 RNN의 기본구조(이기창, 2019)



〈그림 3〉 LSTM 블록 구조(Olah, 2015)

2.3.2 Bidirectional LSTM(BiLSTM)

양방향 장단기 기억 네트워크 신경망인 BiLSTM은 LSTM신경망의 성능을 개선한 알고리즘이다. Schuster and Paliwal(1997)은 일반 RNN의 한계를 극복하기 위해 특정 시간대의 과거와 미래에 이용 가능한 모든 입력 정보를 사용하여 훈련할 수 있는 양방향 반복 신경 네트워크인 Bidirectinal RNN (BRNN)을 제안하였다. 정규 RNN의 상태 뉴런을 양의 시간 방향(Forward states)을 담당하는 부분과 음의 시간 방향(Backward states)을 담당하는 부분으로 나누는 방법으로 이를 통해 RNN의 성능을 향상하였다. 이러한 BRNN 구조를 LSTM에 적용하여 LSTM 모델의 성능을 개선시킨 모델이 BiLSTM이다. 다시 말해, LSTM은 이전 단계의 정보를 메모리에 가지고 있기 때문에 순차적인 시계열(Forward) 예측에 적합하다. 하지만 BiLSTM은 <그림 4>에서 보는 바와 같이, 두 개의 다른 LSTM 네트워크 메모리를 통해 전진 방향(Forward)과 후진 방향(Backward) 모든 시간의 단계에 입력 시퀀스를 최대한 활용하여 훈련함으로써 과거 상황뿐만 아니라



〈그림 4〉 BiLSTM네트워크의 기본구조(Yildirim, 2018)

미래 상황까지 양방향의 정보를 활용하여 과거 상황만 반영할 때의 정보 치우침을 보완한다.

2.3.3 기존연구와의 차이점

기존 한국어 뉴스의 가짜 뉴스 탐지관련 연구는 주로 토픽 모델링(Topic Modeling)으로 각 문서에 포함된 용어의 출현 빈도에 기반하여 유사 문서를 그룹화하고, 각 그룹의 대표 용어들을 추출함으로써 해당 그룹의 토픽 키워드 집합을 제시하는 방식으로 이루어진다. 또한 분류 알고리즘에 있어 SVM, 신경망(Neural Network) 등의 기계학습을 적용하였다(윤태욱, 2018; 현윤진, 김남규, 2018).

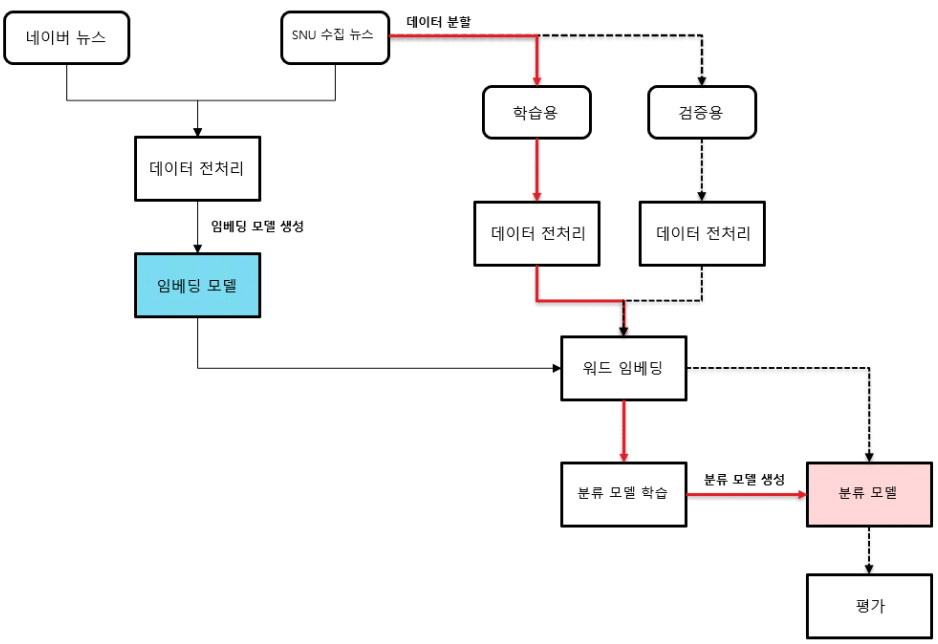
본 연구는 기존 연구에 비해 다음과 같은 차이점이 있다.

첫째, 뉴스 분류를 위해 기존 연구보다 더 많은 실제 뉴스를 수집하였으며, 뉴스의 형식적인 부분보다는 내용상의 의미를 파악하기 위해 자연어의 의미를 내포한 벡터로 표현이 가능하여 자연어 처리에서 많이 활용되고 있는 워드 임베딩 모델(Word2Vec, Fasttext)을 본 연구의 목적에 맞게 실제 한국어 뉴스들을 통해 생성하였다. 둘째, 기존 연구에선 단일 메타 데이터를 사용한데 비해 본 연구에서는 다양한 메타정보를 발굴하여 분류에 반영하였다. 셋째, 메타정보를 개별 독립변수로 사용하지 않고 뉴스 내용과 하나로 합쳐서 임베딩을 하여 메타정보와 뉴스 내용을 하나의 독립변수로 사용하였다. 넷째, 자연어의 토큰(단어)들의 순차적인 의미를 학습하는데 특화된 기법인 LSTM과 LSTM을 양방향(정방향, 역방향)으로 2번 학습하도록 하는 Bi-LSTM 알고리즘, 그리고 워드 임베딩 모델, 메타정보 등을 다양하게 조합하여 최적의 조합을 찾고자 하였다.

Ⅲ. 연구방법

3.1 연구 프레임워크

본 연구의 연구 프레임워크는 다음의 <그림 5>



〈그림 5〉 분류모델 구성

와 같다.

먼저 수집한 모든 뉴스에 대해 형태소 분석과 불용어 처리 등 데이터 전처리 과정을 거쳐 워드 임베딩 모델을 만든다. 그 후, 진위 여부를 분류하고자 하는 Label 정보가 포함된 뉴스를 학습용과 검증용으로 분리하고 데이터 전처리 후 앞서 만든 워드 임베딩 모델을 통해 벡터화 하고, 벡터화된 뉴스를 딥러닝 알고리즘을 통해 진위 여부를 분류하는 모델을 개발하였다.

3.2 분석데이터 및 전처리

3.2.1 데이터 수집

본 연구에서는 뉴스 자료를 두 가지로 나누어 수집하였다. 먼저 임베딩 모델을 만들기 위해서 Label 정보가 없는 뉴스 기사를 네이버 뉴스에서 수집하였다. 수집은 크롤링 프로그램을 코딩하여 수행하였다. 크롤링(Crawling)은 인터넷상에서 존재하는 컨텐츠를 파이썬과 같은 프로그램 언어를 통해 수집을 원하는 컨텐츠를 추출하는 일련의 과정이다. 즉, 특정 웹페이지에 request를 보내고 그 결과를 html형식으로 받은 후, 파서(Parser)를 통해 전달 받은 html에서 필요한 정보를 추출하고 그 결과를 데이터베이스에 저장한다. 본 연구에서는 크롤링을 위해 파이썬의 BeautifulSoup, urllib 라이브러리를 이용하였다. 뉴스 게시 기간은 2020년 1월부터 2020년 8월까지로 키워드 '과학', '국제', '정부', '경제', '국방', '부동산', '정치', '코로나'에 대한 기사 약 30,000건을 수집하였다. 수집한 내용은 '연도', '언론사', '제목', '기사 내용'이며, 중복된 내용의 뉴스 및 사용이 불가한 뉴스를 제외하고 최종적으로 15,129건을 분석에 사용하였다.

다음으로는 뉴스의 진위여부를 분류하기 위한 Label정보가 포함된 뉴스를 수집하였다. 해당 뉴스는 서울대학교 언론정보연구소의 팩트체크 사이트에서 수집하였다. 해당 사이트는 언론사들이 검증한 공적 관심사를 국민에게 알리기 위해 서울대학교 언론정보연구소가 운영하는 정보 서비스이다. 해당 사이트의 검증 대상은 다음과 같으며, 검증 결과는 '전혀 사실 아님', '대체로 사실 아님', '절반의 사실', '대체로 사실', '사실', '판단 유보'의 총 6개의 배지로 표현된다.

① 공직자, 정치인이나 공직자 (예비)후보들이 토론, 연설, 인터뷰, 보도자료 등의 형식으로 발언한 내용의 사실 여부
② 이들 집단과 관련해 언론사의 기사나 소셜미디어 등을 통해 대중에게 회자되는 사실적 진술의 사실성
③ 그 외의 경제, 과학, IT, 사회, 문화 등 제반 분야에서 정확한 사실 검증이 필요하다고 보이는 공적 사안 전반

Label이 있는 뉴스로 수집된 대상은 '전혀 사실 아님', '대체로 사실 아님' 배지(Badge)를 받은 대상(뉴스, SNS 등)과 해당 정보에 대한 검증 뉴스로 총 417건의 뉴스 또는 검증 대상 정보를 해당 사이트에서 직접 수집을 하였다.

3.2.2 데이터 전처리

뉴스는 직접적으로 분석이 가능한 숫자로 표시된 일반적인 정형 데이터와는 달리 다양한 형태의 변이형 자료를 포함하는 비정형 데이터(텍스트)이다. 따라서, 비정형 데이터인 텍스트를 분석이 가능한 형태로 가공하고 정리하는 작업을 반드시 수행해야 하며, 이 과정을 텍스트 전처리라 한다. 전처리 과정에서는 숫자나 문장부호 제거, 오탈자 교정, 대/소문자 통일, 불용어 제거 등의 다양한 작업이 이루어진다(길호현, 2018).

본 연구에서는 전처리 작업 과정에서 문장부호 제거, 어근 동일화, 불용어 처리를 수행하였다. 문장부호 제거, 어근 동일화는 KONLPY konlpy.org/en/latest/의 Twitter 형태소 분석을 통해 수행하였고, 불용어 제거는 한국어 불용어 사전으로 많이 활용되는 자료와 본 연구에서 수집한 뉴스 내용 중 불필요한 문구(예: 구독하기, 제보하기)를 취합하여 불용어 사전을 새로 구축한 후, 이를 이

용하여 제거하였다. 불용어 제거 시 불용어 대상 글자의 길이에 따라 두 가지 방법으로 나눠서 처리하였다.

첫 번째, 불용어 글자의 길이가 세 글자 이하일 경우 뉴스 내용에서 직접 제거할 때 '인'을 삭제하면 '인터넷'의 '인'이 삭제되는 등 의도하지 않은 대상까지 삭제가 되어 형태소 분석을 한 다음 형태소 단위로 비교하여 제거하였다.

두 번째, 불용어 글자의 길이가 4글자 이상인 경우에는 위와 반대로 형태소 분석 후 비교를 하게 되면 '다시 말하면'이 '다시/Noun' + '말/Noun' + '하다/Verb'로 변환이 되어 글자의 길이도 짧아지고, 형태도 변경이 되는 문제가 있어 뉴스문장에서 직접 삭제하였다.

3.2.3 데이터 설명

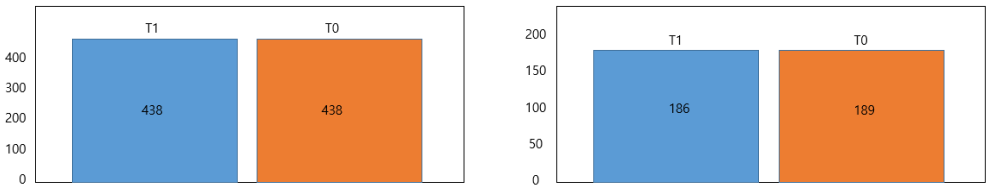
본 연구에서는 서울대학교 팩트체크 사이트에서 진위 여부를 분류할 수 있는 417건의 뉴스와 해당 뉴스의 메타정보를 수집하였으며, 이를 기반으로 뉴스의 진위여부를 가리는데 필요한 메타정보를 직접 생성하였다. 팩트체크 사이트에서 수집한 정보는 뉴스 제목, 본문, 메타 정보(주제, 소주제, 주장/검증 매체)이며, 연구를 위해 생성한 메타 정보는 뉴스이 '대상'과 뉴스의 '주장/검증 주체' 데이터이다. 직접 생성한 메타정보 중 '대상'은 뉴스에서 중점적으로 다루어지는 주제로서 '소주제'가 같은 코로나 바이러스이더라도 정부기관과 관련된 내용이면 '정부', 의학지식에 대한 내용이면 '의학지식' 등으로 분류하였다. '주장/검증 주체'는 해당 뉴스를 작성 또는 배포한 주체로 같은 SNS에 올라온 주장이라도 일반인이 올린 주장이면 '일반인', 정치인이 올린 주장이면 '정치인'으로 분류하였다. <표 3>은 수집된 뉴스와 메타 데이터의 정의를 보여준다.

수집된 뉴스는 학습용과 검증용으로 분리하였다. 전체 417건(연관ID 기준 201건) 중 학습용은 292건, 검증용은 125건으로 7:3비율로 분리하였다. 수집된 뉴스의 데이터 건수가 많지 않아 실험을 위해 over-sampling을 통해 학습용, 검증용 뉴스를 각각 3배수로 증가시켜 학습용 876건, 검증용 375건으로 분석하였다.

데이터 전처리의 마지막 작업으로, 학습/검증용 뉴스의 Label 비율은 데이터 균형화 작업을 통해 <그림 6>과 같이 각각 438(가짜):438(진짜), 186(가짜):186(진짜)으로 동일하게 맞췄다.

〈표 3〉 수집된 뉴스와 메타 데이터 정의

변수명	변수 설명	수집출처
연관ID	진짜와 가짜를 연결한 ID	직접 생성
ROW ID	ROW당 ID	직접 생성
주제	뉴스의 주제	SNU
소주제	뉴스의 상세주제	SNU
내용	뉴스의 제목	SNU
상세내용	뉴스의 본문	SNU
대상	뉴스의 중심대상이 되는 주체	직접 생성
주장/검증 주체	뉴스의 중심내용을 주장 또는 검증한 주체	직접 생성
주장/검증 매체	뉴스를 주장 또는 검증한 매체	SNU
Label	진위 여부	SNU



〈그림 6〉 학습용, 검증용 뉴스 비율

3.3 분류모델생성

3.3.1 개발환경

본 연구는 Google Colaboratory을 사용하여 Tensorflow 1.15.2 버전으로 구현하였다. Google Colaboratory는 구글에서 교육과 과학 연구를 목적으로 개발한 도구로, 2017년에 무료로 공개하였다. 줄여서 'Colab'이라고도 부르며, 특별한 개발 환경이 PC에 구축되어 있지 않아도 브라우저에서 Python을 작성하고 실행할 수 있다. Tensorflow는 구글(Google)사에서 개발한 기계 학습(machine learning) 엔진, 검색, 음성 인식, 번역 등의 구글 앱에 사용되는 기계 학습용 엔진으로, 2015년에 공개 소스 소프트웨어(Open Source Software)로 전환되었다. Tensorflow는 C++ 언어로 작성되었고, 파이썬(Python) 응용 프로그래밍 인터페이스(API)를 제공한다.

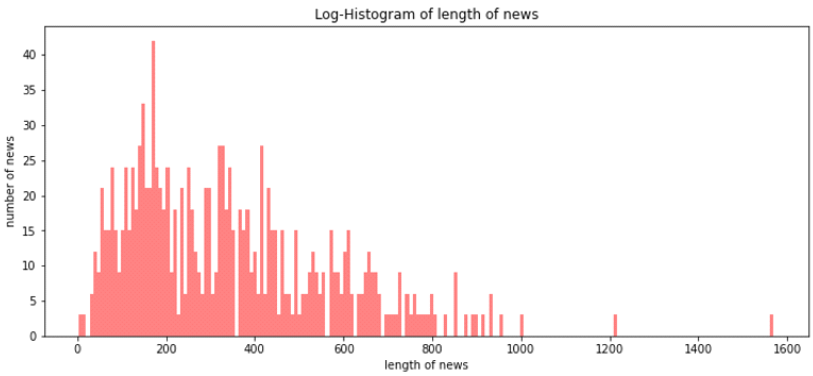
3.3.2 실험설계

3.3.2.1 토큰화(Tokenization)

본 연구에서는 수집된 뉴스를 형태소 단위로 분리하는 토큰화 작업을 수행하기 위해 KONLPY 패키지의 Twitter를 사용하였다. 토큰화 과정에서 불용어 길이별로 두 가지의 불용어 사전을 적용하

〈표 4〉 팩트체크 사이트에서 수집한 뉴스의 토큰화 후 분포 비교

형태소 분석 후 분포 비교	값(건, %)	비고
전체 형태소 개수	12,969	
희귀 단어 개수	5,065	출현빈도가 4회 이하
전체 형태소에서 희귀 단어 비율	39.05	
전체 등장빈도에서 희귀단어 등장 비율	3.54	



〈그림 7〉 뉴스의 토큰 개수 분포

여 불필요한 요소는 제거하였다. 분류를 위한 뉴스의 경우 <표 4>와 같이 전체 형태소 중 출현 빈도가 4회 이하인 단어가 39%를 차지하였고, 해당 단어가 등장한 빈도는 전체의 약 4%로 확인되었다.

또한 분류를 위해 수집한 뉴스의 토큰화 결과 가장 많은 토큰을 가진 뉴스의 토큰 개수는 1,569개이고, 평균 토큰 수는 342.61개로 나타났다. 뉴스별 토큰 개수의 분포는 <그림 7>과 같다.

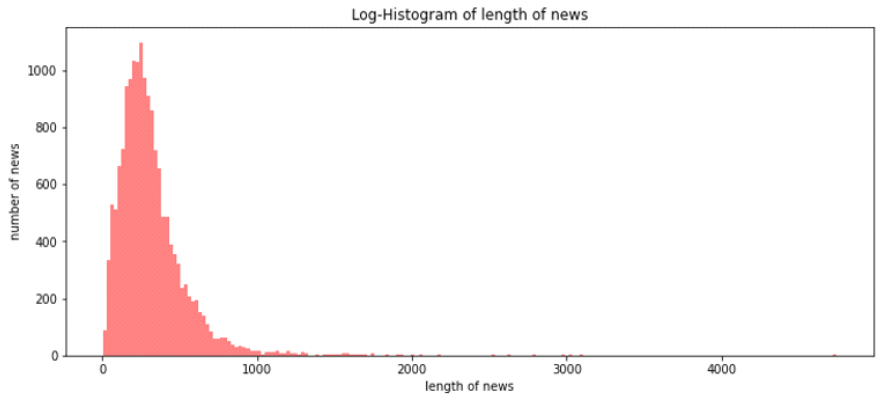
3.3.2.2 워드 임베딩 모델(Word Embedding Model)

본 연구에서는 형태소 단위로 토큰화된 뉴스를 벡터(Vector)로 변환하기 위해 Word2Vec과 FastText를 각각 적용하여 워드 임베딩 모델을 생성하였다. 임베딩 모델은 학습 시 주변 단어와의 관계를 통해 단어에 의미를 부여하게 된다. 그러므로 뉴스기사를 학습하여 만들어진 임베딩 모델을 가짜뉴스 분류에 활용하는 것이 적합하다고 판단하여 네이버 뉴스에서 크롤링한 뉴스와 팩트체크 사이트에서 직접 수집한 뉴스 총 16,380개를 사용하였고 불용어 사전을 적용하였으며, KONLPY의 Twitter를 통해 형태소 분석을 수행하였다. 가장 많은 토큰을 가지는 뉴스의 토큰 개수는 4,747개이며, 평균 토큰 수는 320개로 나왔다. 뉴스별 토큰 개수의 분포는 <그림 8>과 같다.

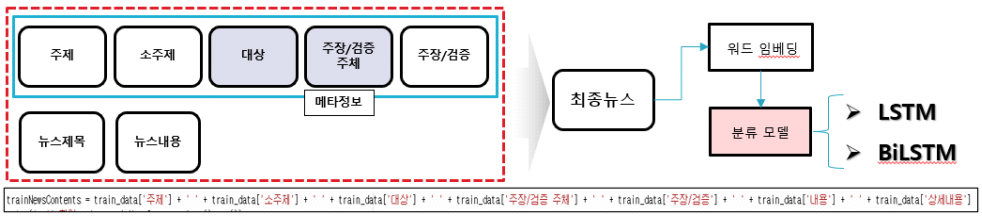
임베딩 모델의 하이퍼 파라미터(Hyper Parameter)의 경우, 차원은 300차원, 윈도우 사이즈는 5개, 최소 횟수는 3회로 설정하였다.

3.3.2.3 분류모델

본 연구에서는 자연어 처리와 텍스트 분류에 많이 사용되고 있는 딥러닝 알고리즘인 LSTM 및 양방향LSTM(BiLSTM)과 임베딩 알고리즘인 Word2vec 및 Fasttext를 사용하여 하이퍼 파라미터 조정에 따른 분류 정확도를 비교 실험하였다. 분류 모델의 구성도는 <그림 9>와 같다.



〈그림 8〉 임베딩 모델을 만들기 위한 뉴스의 토큰 개수 분포



〈그림 9〉 분류 모델 구성도

3.3.2.4 모델 성과 검증지표

본 연구에서는 개발된 모델의 분류 정확도를 비교하기 위해 하이퍼 파라미터 및 검증지표 측정표를 구성하였으며, 검증지표로 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), 그리고 정밀도와 재현율의 조화평균인 F1-Score를 사용하였다.

Ⅳ. 실험결과

4.1 실험결과

본 연구에서는 Word2Vec과 Fasttext로 워드임베딩 모델을 만들어 각각 LSTM과 BiLSTM 알고리즘에 적용하여 실험을 진행하였다. 1차 실험에서는 LSTM과 BiLSTM 각각에 대해 batch size를 조정하면서 두 분류모델의 정확도를 비교하였다.

각각의 분류 알고리즘을 Word2Vec과 Fasttext로 2회씩 총 4회의 실험을 진행하였다. LSTM의 평균 정확도는 0.58로 나왔고, BiLSTM의 평균 정확도는 0.76으로 나와, BiLSTM이 LSTM보다 본 연구에서 더 좋은 분류 정확도를 보였다. LSTM의 가장 높은 정확도를 보인 차수2 + Word2Vec의 경우 <표 5>의 검증결과를 보였고, BiLSTM의 가장 높은 정확도를 보인 차수2 + Fasttext의 경우 <표 6>과 같은 검증결과를 보였다.

1차 실험에서 BiLSTM이 LSTM에 비해 본 연구

〈표 5〉 LSTM과 BiLSTM의 정확도 비교 결과

분류 기법	임베딩 모델	실험 차수	하이퍼 파라미터				검증지표
			Batch size	Learing rate	Node	Epoch	Accuracy
LSTM	word2vec	차수 1	16	0.001	256	8	0.57
	Fasttext						0.58
LSTM	word2vec	차수 2	32	0.001	256	8	0.62
	Fasttext						0.54
평균							0.58
BiLSTM	word2vec	차수 1	16	0.001	256	8	0.76
	Fasttext						0.76
BiLSTM	word2vec	차수 2	32	0.001	256	8	0.71
	Fasttext						0.80
평균							0.76

〈표 6〉 1차 실험 차수2 LSTM+Word2Vec

구분		Precision	Recall	F1-Score	Accuracy
Target	진짜	0.71	0.40	0.51	0.62
	가짜	0.58	0.84	0.68	

〈표 7〉 1차 실험 차수2BiLSTM+Fasttext

구분		Precision	Recall	F1-score	Accuracy
Target	진짜	0.8	0.8	0.8	0.80
	가짜	0.8	0.8	0.8	

〈표 8〉 2차 실험 구성

임베딩 모델	Batch size	Learning rate	Node	Epoch
Word2vec	16 / 32	0.001 / 0.01	256 / 512	8 / 16
Fasttext				

〈표 9〉 BiLSTM+Word2Vec 실험 결과

실험차수	Batch size	Learning rate	Node	Epoch	Accuracy
차수1	16	0.001	256	8	0.76
차수2	16	0.001	256	16	0.77
차수3	16	0.001	512	8	0.70
차수4	16	0.001	512	16	0.72
차수5	16	0.01	256	8	0.77
차수6	16	0.01	256	16	0.76
차수7	16	0.01	512	8	0.84
차수8	16	0.01	512	16	0.72
차수9	32	0.001	256	8	0.71
차수10	32	0.001	256	16	0.72
차수11	32	0.001	512	8	0.71
차수12	32	0.001	512	16	0.76
차수13	32	0.01	256	8	0.78
차수14	32	0.01	256	16	0.80
차수15	32	0.01	512	8	0.78
차수16	32	0.01	512	16	0.77
평균					0.75

에서 더 좋은 분류 정확도를 보였기에, 2차 실험에서는 BiLSTM 기법을 중심으로 임베딩 모델과 하이퍼 파라미터를 변경하면서 최적의 조합을 찾고자 하였다.

<표 8>과 같이 각각의 임베딩 모델별로 batch size, learning rate, node, epoch를 각각 두 개의 경우로 나누어 하이퍼 파라미터셋을 구성하였다. 각 임베딩 모델별로 실험을 수행하여, 총 32번의 실험을 수행하였다.

<표 9>는 Word2Vec으로 실험한 결과를 보여준다. 차수 7의 정확도가 0.84로 가장 높았고, 차수 3의 정확도가 0.70으로 가장 낮았다. 정확도가 0.75 이하가 6개, 0.80 이하가 9개, 0.85 이하가 1개로 전체 평균은 0.75가 나왔다.

<표 10>은 Fasttext로 실험한 결과이다. 차수 9의 정확도가 0.80으로 가장 높았고, 차수 11의 정확도가 0.73으로 가장 낮았다. 정확도가 0.75 이하가 5개, 0.80 이하가 11개로 전체 평균은 0.77이 나왔다. Word2Vec에서 가장 높은 정확도를 보인 차수 7의 검증 결과와 Fasttext에서 가장 높은 정확도를 보인 차수 9의 검증 결과를 비교하면 <표 11>과 같다.

두 임베딩 모델의 정확도는 Word2Vec에서 가장 높은 0.84를 보였고, Fasttext는 0.80의 최고 정확도를 보여 단일 실험 차수에서는 Word2Vec이 0.04 높았다. Paired t-test 결과 이러한 정확도의 차이는 통계적으로 유의한 것으로 나왔다(p<0.05). <그림 10>은 Word2Vec과 Fasttext의 차수별 정확

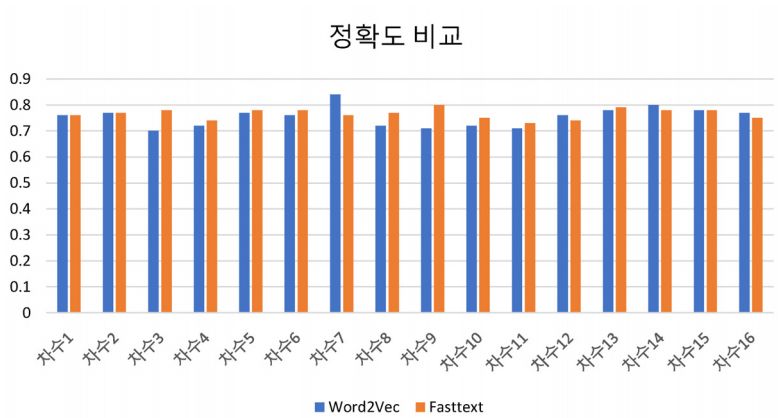
〈표 10〉 BiLSTM+Fasttext 실험 결과

실험차수	Batch size	Learning rate	Node	Epoch	Accuracy
차수1	16	0.001	256	8	0.76
차수2	16	0.001	256	16	0.77
차수3	16	0.001	512	8	0.78
차수4	16	0.001	512	16	0.74
차수5	16	0.01	256	8	0.78
차수6	16	0.01	256	16	0.78
차수7	16	0.01	512	8	0.76
차수8	16	0.01	512	16	0.77
차수9	32	0.001	256	8	0.80
차수10	32	0.001	256	16	0.75
차수11	32	0.001	512	8	0.73
차수12	32	0.001	512	16	0.74
차수13	32	0.01	256	8	0.79
차수14	32	0.01	256	16	0.78
차수15	32	0.01	512	8	0.78
차수16	32	0.01	512	16	0.75
평균					0.77

〈표 11〉 Word2Vec 실험 차수7, Fasttext 실험 차수9 결과 비교

임베딩 모델		Precision	Recall	F1-Score	Accuracy	Accuracy 차이
Word2vec	진짜	0.82	0.88	0.85	0.84	0.04**
	가짜	0.87	0.80	0.83		
Fastext	진짜	0.80	0.80	0.80	0.80	
	가짜	0.80	0.80	0.80		

** $p < 0.05$

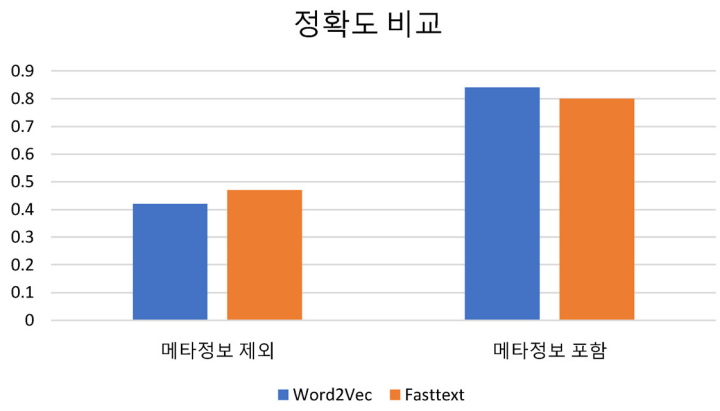


〈그림 10〉 Word2Vec과 Fasttext의 차수 별 정확도 비교

〈표 12〉 메타 정보 사용 여부에 따른 정확도 비교

임베딩 모델	Accuracy		차이
	메타정보 제외	메타정보 포함	
Word2vec	0.42	0.84	0.42***
Fastext	0.47	0.80	0.33***

*** $p<0.01$



〈그림 11〉 메타 정보 사용 여부에 따른 정확도 비교

도를 비교한 것으로, Fasttext의 평균 정확도가 0.77로 Word2Vec의 평균 정확도 0.75보다 높게 나타나 BiLSTM과 Fasttext의 조합이 평균적으로는 가장 좋은 분류 정확도를 보였다.

최종 3차 실험에서는 메타정보의 활용이 모델의 분류 정확도에 영향을 미치는 정도를 확인하고자 2차 실험에서 각 임베딩 모델별로 가장 좋은 정확도를 보였던 차수의 하이퍼 파라미터를 활용하여 메타정보 없이 뉴스제목과 뉴스내용만으로 추가 실험을 수행하였다.

3차 실험 결과 <표 12>와 같이, 메타정보를 모델에서 제외할 경우 모델에 포함하는 경우에 비해서 정확도가 Word2Vec은 0.42, Fasttext는 0.33 낮게 나왔으며, 이러한 정확도의 차이는 paired t-test 결과 통계적으로 유의한 것으로 나왔다. 이를 통해 메타정보가 뉴스의 진위분류에 많은 영향을 미치는 것을 확인할 수 있다.

Ⅴ. 결 론

5.1 요약 및 시사점

"진실은 너무 교활해서 붙잡기 힘들다."라고 셰익스피어가 말했다. 그만큼 가짜 뉴스는 오래전부터 우리 곁에 있었던 그렇게 특별한 문제는 아니다. 하지만 정보화시대로 접어들면서 정보의 입수처가 다양해지고, 양은 많아졌으며, 소비의 속도도 빨라졌다. 특히 인터넷상에서 우리가 접하는 대다수의 뉴스는 검증되는 시간보다 전파되는 시점이 더 빨라져 갈수록 진실 여부를 고민할 시간이 짧아져 가는 상황이다. 따라서, 가짜뉴스에 대한 검증 정확도와 함께 검증의 속도도 함께 필요한 실정이다.

본 연구는 이러한 측면에서 가짜 뉴스를 빠르게 판별하기 위해 뉴스의 진위를 분류할 수 있는

딥러닝 기반의 모델을 개발하는 것을 목적으로 하였다. 이를 위해 Word2vec과 Fasttext와 같은 워드 임베딩 모델과 LSTM과 BiLSTM과 같은 딥러닝 기법을 사용하여, 여러 차수의 실험을 진행하였으며, 그 결과 최고 84%의 정확도를 보이는 BiLSTM-Word2vec기반의 분류 모델을 개발하였다.

본 연구는 다음과 같은 점에서 학문적 시사점이 있다.

첫째, 기존의 가짜뉴스 판별을 위한 연구에서는 뉴스 기사 이외에 '언론사'라는 제한적인 메타정보를 사용하였거나(윤태욱, 2018), 트위터 정보를 활용하였지만(Nikam, 2020; 현윤진 등, 2018), 본 연구에서는 선행연구에서 사용하지 않았던 다양한 메타정보를 발굴한 후 모델에 포함하여 뉴스 자체의 컨텐츠 외 해당 뉴스를 설명할 수 있는 메타 정보들의 활용을 통해 보다 정확도가 높은 분류 모델을 개발하였으며, 이를 통해 메타 정보를 함께 활용하는 것이 모델 성능에 좋은 영향을 미친다는 것을 검증하였다. 둘째, 워드 임베딩과 분류 모델 개발 시 인공적으로 생성 및 조작된 뉴스가 아닌 공신력 높은 기관에서 진위가 판단된 실제 뉴스를 수집하여 분석에 사용함으로써, 분석 결과의 신뢰도를 높였다. 셋째, 메타정보를 단순히 개별 독립변수로 사용한 것이 아니라, 뉴스의 내용과 모든 메타정보를 하나로 합쳐 임베딩함으로써, 메타정보와 뉴스의 내용을 하나의 독립변수로 사용하였고 이를 통해 보다 종합적인 분석이 가능하도록 하였다. 넷째, 기존 연구가 다루지 않았던 워드 임베딩 기술과 딥러닝 기법을 활용하여 분류 정확도를 비교하여 보다 적합한 임베딩 모델과 딥러닝 기법의 조합을 제시하였다.

또한, 본 연구는 다음과 같은 점에서 실무적 시사점이 있다.

첫째, 기존 연구보다 더 정확하게 뉴스의 진위여부를 판단할 수 있게 되어, 가짜뉴스 판별 서비스의 질을 높였고, 이를 통해 네이버와 다음 등 국내 포털 서비스 업체에서 사용자들에게 뉴스를 제공할 경우 해당 뉴스의 진위여부를 더욱 정확하게 검증하는데 활용이 가능할 것으로 보인다. 둘째, 메신저, SNS 서비스 제공 업체에서는 사용자들이 전파하는 뉴스정보들의 진위 여부를 빠르고 정확하게 판단하여 정보 수신자들에게 알려줌으로써 수신자들이 가짜 뉴스에 현혹되지 않도록 하고, 가짜 뉴스 전파를 미연에 방지하여 가짜 뉴스가 유행되지 않도록 하는데 도움이 될 것으로 보인다. 셋째, 또한 일반인들도 본인이 읽고 있는 뉴스의 진위여부를 알고자 할 경우 보다 쉽게 확인이 가능할 것으로 기대된다. 넷째, 뉴스와 관련된 메타 정보의 중요성을 보여줌으로써, 가짜뉴스 예측 서비스를 제공하는 업체들에게 메타정보의 수집 및 발굴에 대한 필요성을 제시하였다.

5.2 한계점 및 향후 연구 방향

본 연구는 다음과 같은 측면에서 한계점을 지닌다.

첫째, 현실적인 진위가 판단된 뉴스 데이터 수집의 어려움으로 인해, 본 연구에서 분석에 사용한 학습 데이터의 수가 적다는 한계점을 갖는다. 하지만 데이터는 시간이 지남에 따라 계속 축적됨으로 추후 보다 많은 데이터를 분석에 활용할 수 있을 것으로 보인다. 둘째, 실험 환경의 한계로 인해 보다 다양한 하이퍼 파라미터에 대한 실험이 이루어지지 못했다. 이러한 점은 향후 추가적인 실험을 통해 보완하고자 한다. 향후 연구에서는 위와 같은 한계점들을 보완하고, 더불어 워드 임베딩 방식도 단어수준에서 문장수준으로 확장한 분류 모델을 개발하여 본 연구의 결과와 비교해 보고자 한다.

참 고 문 헌

[1] 길호현, "텍스트마이닝을 위한 한국어 불용어 목록 연구", *우리말글*, 제78집, 2018, pp. 1-25.

[2] 김유향, "미 대선 시기 가짜뉴스(Fake News) 관련 논란과 의미", *국회입법조사처 이슈와*

논점, 제1242호, 2016.
[3] 뉴스퀘어, "가짜 뉴스가 판 치는 세상", 2017.
[4] 버즈피드 뉴스, Available at https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook.
[5] 성욱제, 정은진, "코로나19와 허위정보: 유형 분석과 대응방안", 연구보고서, 정보통신정책연구원, 2020.
[6] 양정애, "일반 시민들이 생각하는 '뉴스'와 '가짜뉴스'", *한국언론진흥재단 Media Issue*, 제5권, 제1호, 2019.
[7] 유은조, 이지현, 박소영, "LSTM 모델을 통한 국문 기사 감성 분류 시스템", *한국정보과학회 학술발표논문집*, 2018, pp. 1949-1951.
[8] 윤영석, 엄태원, 안재영, 이현우, 허재두, "페이크 뉴스 탐지 기술 동향과 시사점", 연구보고서, 정보통신기술진흥센터, 2017.
[9] 윤태욱, *토픽모델링과 SVM을 이용한 한국어 가짜뉴스 탐지 시스템*(석사학위논문), 국민대학교 비즈니스IT전문대학원, 2018.
[10] 이기창, 한국어 임베딩, 에이콘출판, 2019.
[11] 조현수, 이상구, "FastText를 적용한 한국어 단어 임베딩", *한국소프트웨어종합학술대회 논문집*, 제12호, 2017, pp. 705-707.
[12] 좌희정, 오동석, 임희석, "자동화기반의 가짜 뉴스 탐지를 위한 연구 분석", *한국융합학회논문지*, 제10권, 제7호, 2019, pp. 15-21.
[13] 주원, 정민, 백다미, "가짜 뉴스(Fake News)의 경제적 비용 추정과 시사점", 연구보고서, 현대경제연구원, 2017.
[14] 진민정, "프랑스 대선보도와 가짜뉴스-언론 · SNS · 정치권 · 교육계 모두 '가짜 뉴스와 전쟁'", *한국언론진흥재단 신문과방송*, 557호, 2017.
[15] 채상희, *한국어 감성분석을 위한 텍스트 임베딩 방법론 연구*(석사학위논문), 서울시립대학교 일반대학원, 2019.
[16] 현윤진, 김남규, "뉴스와 소셜 데이터를 활용한 텍스트 분석 기반 가짜 뉴스 탐지 방안", *한국전자거래학회지*, 제23권, 제4호, 2018, pp. 19-39.
[17] 황용석, "가짜뉴스 개념 정의의 문제-형식과 내용 의도적으로 속일 때 '가짜뉴스'", *한국언론진흥재단 신문과 방송*, 2017.
[18] Allcott, H. and M. Gentzkow, "Social media and fake news in the 2016 election", *Journal of Economic Perspectives*, Vol.31, No.2, 2017, pp. 211-36.
[19] Hochreiter, S. and J. Schmidhuber, "Long short-term memory", *Journal of Neural Computation*, Vol.9, No.8, 1997, pp. 1735-1780.
[20] Nikam, S. S. and R. Dalvi, "Machine learning algorithm based model for classification of fake news on Twitter", *Fourth International Conference on I-SMAC*, 2020.
[21] Olah, C., "Understanding LSTM Networks", 2015, Available at http://colah.github.io/posts/2015-08-Understanding-LSTMs/.
[22] Schuster, M. and K. K. Paliwal, "Bidirectional recurrent neural networks", *IEEE Transactions on Signal Processing*, Vol.45, 1997, pp. 2673-2681.
[23] Williams, R. J., G. E. Hinton, and D. E. Rumelhart, "Learning representations by back-propagating errors", *Nature*, Vol.323 , No.6088, 1986, pp. 533-536.
[24] Yildirim, O., "A novel wavelet sequences based on deep bidirectional LSTM network model for ECG signal classification", *Computer in Biology and Medicine*, Vol.96, No.1, 2018, pp. 189-202.
[25] Yuan, P. Y., A. M. Du, and C. Wang, "Using Word2vec to match knowledge points and test questions: A case study", *IEEE 2nd International Conference on Computer Science and Educational Informatization (CSEI)*, 2020, pp. 272-276.

Information Systems Review
Volume 23 Number 4
November 2021

Development of a Fake News Detection Model Using Text Mining and Deep Learning Algorithms

Dong-Hoon Lim* · Gunwoo Kim** · Keunho Choi***

Abstract

Fake news isexpanded and reproduced rapidly regardless of their authenticity by the characteristics of modern society, called the information age. Assuming that 1% of all news are fake news, the amount of economic costs is reported to about 30 trillion Korean won. This shows that the fake news isvery important social and economic issue. Therefore, this study aims to develop an automated detection model to quickly and accurately verify the authenticity of the news. To this end, this study crawled the news data whose authenticity is verified, and developed fake news prediction models using word embedding (Word2Vec, Fasttext) and deep learning algorithms (LSTM, BiLSTM). Experimental results show that the prediction model using BiLSTM with Word2Vec achieved the best accuracy of 84%.

Keywords: Fake News, Korean News, Natural Language Processing, Deep Learning, Text Mining

* Deputy General Manager, Data World
** Professor, Department of Business Administration, Hanbat National University
*** Corresponding Author, Assistant Professor, Department of Business Administration, Hanbat National University

◐ 저 자 소 개 ◑

임 동 훈 (dhcrom@naver.com)

한밭대학교에서 경영학 석사학위를 수여하였으며, 현재 ㈜데이터월드에 재직 중이다. 관세청 등 국가공공기관 관련 시스템 구축 및 유지 사업에 참여하고 있으며, 주요 관심분야는 머신러닝, 딥러닝, 자연어처리, 연관 규칙 등이다.

김 건 우 (gkim@hanbat.ac.kr)

고려대학교에서 경영학 박사학위를 수여하였으며, 현재 한밭대학교 융합경영학과에서 교수로 재직 중이다. 주요 관심분야는 비즈니스 온톨로지 모델, 빅데이터 분석, 핀테크 기술 및 전략 등이다.

최 근 호 (keunho@hanbat.ac.kr)

고려대학교에서 경영학 박사학위를 수여하였으며, 현재 한밭대학교 융합경영학과에서 조교수로 재직 중이다. 주요 관심분야는 추천시스템, 의료 빅데이터 분석, 딥러닝, 머신러닝 등이다.

논문접수일 : 2021년 08월 03일 **게재확정일 : 2021년 10월 16일**
1차 수정일 : 2021년 09월 14일