KIPS Trans. Softw. and Data Eng.
Vol.10, No.12 pp.569~578
ISSN: 2287-5905 (Print), ISSN: 2734-0503 (Online)

https://doi.org/10.3745/KTSDE.2021.10.12.569

A Study on Automated Fake News Detection Using Verification Articles

Yoon-Jin Han† · Geun-Hyung Kim††

ABSTRACT

Thanks to web development today, we can easily access online news via various media. As much as it is easy to access online news, we often face fake news pretending to be true. As fake news items have become a global problem, fact-checking services are provided domestically, too. However, these are based on expert-based manual detection, and research to provide technologies that automate the detection of fake news is being actively conducted. As for the existing research, detection is made available based on contextual characteristics of an article and the comparison of a title and the main article. However, there is a limit to such an attempt making detection difficult when manipulation precision has become high. Therefore, this study suggests using a verifying article to decide whether a news item is genuine or not to be affected by article manipulation. Also, to improve the precision of fake news detection, the study added a process to summarize a subject article and a verifying article through the summarization model. In order to verify the suggested algorithm, this study conducted verification for summarization method of documents, verification for search method of verification articles, and verification for the precision of fake news detection in the finally suggested algorithm. The algorithm suggested in this study can be helpful to identify the truth of an article before it is applied to media sources and made available online via various media sources.

Keywords : Artificial Intelligence, Fake News Detection, Fake News Data, KoBERT, Text Summarization

검증 자료를 활용한 가짜뉴스 탐지 자동화 연구

한 윤 진† · 김 근 형††

요 약

오늘날 웹의 발전으로 우리는 각종 언론 매체를 통해 온라인 기사를 쉽게 접하게 된다. 온라인 기사를 쉽게 접할 수 있게 된 만큼 거짓 정보를 진실로 위장한 가짜뉴스 또한 빈번하게 찾아볼 수 있다. 가짜뉴스가 전 세계적으로 대두되면서 국내에서도 가짜뉴스를 탐지하기 위한 팩트 체크 서비스가 제공되고 있으나, 이는 전문가 기반의 수동 탐지 방법을 기반으로 하며 가짜뉴스 탐지를 자동화하는 기술에 대한 연구가 계속해서 활발하게 이루어지고 있다. 기존 연구는 기사 작성에 사용된 문맥의 특성이나, 기사 제목과 기사 본문의 내용 비교를 통한 탐지 방법이 가장 많이 사용되고 있으나, 이러한 시도는 조작의 정밀도가 높아졌을 때 탐지가 어려워질 수 있다는 한계를 가진다. 따라서 본 논문에서는 기사 조작의 발달에 따른 영향을 받지 않기 위하여 기사의 진위 여부를 판단할 수 있는 검증기사를 함께 사용하는 방법을 제안한다. 또한 가짜뉴스 탐지 정확도를 개선시킬 수 있도록 실험에 사용되는 기사와 검증기사를 문서 요약 모델을 통해 요약하는 과정을 추가했다. 본 논문에서는 제안 알고리즘을 검증하기 위해 문서 요약 기법 검증, 검증기사 검색 기법 검증, 그리고 최종적인 제안 알고리즘의 가짜뉴스 탐지 정확도 검증을 진행하였다. 본 연구에서 제안한 알고리즘은 다양한 언론 매체에 적용하여 기사가 온라인으로 확산되기 이전에 진위 여부를 판단하는 방법으로 유용하게 사용될 수 있다.

키워드 : 인공지능, 가짜뉴스 탐지, 가짜뉴스 데이터, KoBERT, 문서 요약

1. 서 론

지난 2016년 미국 대통령 선거에서 가짜뉴스(Fake News) 문제가 많은 영향을 미치면서 전 세계적으로 사회적 이슈가 되었다. 미국 대통령 선거 기간 동안 다양한 매체들의 진짜뉴스보다 더 이슈가 된 가짜뉴스가 많은 반응을 이끌어내어 정치적으로 피해를 입혔다. 가짜뉴스는 정치뿐만 아니라, 경제, 국제, 문화 등 다양한 주제의 뉴스에서도 찾아볼 수 있다. 웹의 발전으로 인해 가짜뉴스는 더욱 쉽게 확산될 수 있으므로 영향력이 갈수록 확대되고 있으며, 진짜뉴스와 비슷한 문맥으로 작성되는 양상으로 여론을 더 불안하고 혼란스럽게 만들고 있다[1,2].

따라서 가짜뉴스가 확산되는 것을 미리 탐지하여 사전에 방지하는 것이 오늘날 중요한 문제가 되었으며, 다양한 가짜뉴스 탐지 방법이 제시되고 있다. 가짜뉴스 탐지 방법은 비기술적 탐지 방법과 기술적 탐지 방법으로 구분된다. 가짜뉴스 탐지를

※ 본 연구는 2021년도 동의대학교 교내연구비에 의해 연구되었음. (202102000001).
† 준 회 원 : 동의대학교 디지털미디어공학과 석사과정
†† 종신회원 : 동의대학교 게임공학전공 교수
Manuscript Received : June 14, 2021
First Revision : July 21, 2021
Accepted : August 18, 2021
* Corresponding Author : Geun-Hyung Kim(geunkim@deu.ac.kr)

위한 비기술적 탐지 방법으로는 전문가 기반 가짜뉴스 탐지 방법과 집단지성 기반 가짜뉴스 탐지 방법이 있다. 해당 방법들은 각각 전문성을 가진 기자와 신뢰성에 의구심을 가진 사용자들이 뉴스의 진위 여부를 판단한다. 비기술적 탐지 방법은 사람이 판단한다는 점에서 검증 과정과 결과가 명료하여 공신력을 가지지만, 오랜 분석 시간이 필요하고 한정된 웹 사이트 내에서만 판단 결과를 확인할 수 있다는 단점을 가진다[2].

가짜뉴스 탐지를 위한 기술적 방법으로는 가짜뉴스 데이터를 기계에 학습시켜 판단할 뉴스의 가짜뉴스 확률을 추정하는 인공지능 기반 가짜뉴스 탐지 방법과 언어학을 기반으로 뉴스에 사용된 단어, 어절, 문장, 맥락 등을 분석하여 내용의 사실성을 검증하는 시맨틱 기반 가짜뉴스 탐지 방법이 있다.

그 중, 인공지능 기반의 기존 연구들은 기사 작성에 사용된 문맥의 특성이나, 기사 제목과 기사 본문의 내용 비교를 통한 탐지 방법이 가장 많이 사용되고 있으나, 이러한 시도는 조작의 정밀도가 높아졌을 때 탐지가 어려워질 수 있다는 한계를 가진다. 기사의 형태를 분석하여 공인된 표준 규격을 기준으로 형태가 올바르지 않는 기사를 탐지하는 방법도 가짜뉴스가 표준 규격에 맞게 작성된다면 가짜뉴스 판별이 어렵다. 즉, 판단하고자 하는 기사 자체만으로는 가짜뉴스를 식별하는데 어려움이 있으니 정확도를 개선시키기 위해선 추가로 활용할 검증 데이터가 필요하다[3-5].

본 논문에서는 기사 조작의 발달에 따른 영향을 받지 않기 위하여 기사의 진위 여부를 판단할 수 있는 검증기사를 함께 사용하는 방법을 제안한다. 또한 문서 요약 모델을 사용하여 기사와 검증기사에 요약을 거쳐 가짜뉴스 탐지 모델이 더욱 효율적으로 학습하는 방안을 제시한다.

또한, 기존의 한국어 가짜뉴스 탐지 연구는 자연어 처리 단계에서 단어의 빈도수를 활용하여 문서의 특성을 파악하는 TF-IDF를 활용한 모델들이 다수였다[6]. 본 논문에서는 문장의 맥락을 표현할 수 없는 TF-IDF와 다르게 대용량 데이터로 사전 학습되어 문맥 파악에 수월한 KoBERT를 사용한다. 학습된 모델을 새로운 데이터로 테스트하기 위해 판단 대상인 기사에서 키워드를 추출한 후, 키워드를 통해 다양한 URL에서 검색된 기사들을 검증기사로 사용할 것이다.

논문의 구성은 다음과 같다. 2장에서 제안 알고리즘과 관련된 연구동향을 정리하고, 3장에서 제안 알고리즘의 방법론과 사용된 데이터를 소개한다. 4장에서 제안 알고리즘을 검증하기 위한 실험을 거쳐 마지막 5장 결론에서는 실험 결과와 함께 더 나은 가짜뉴스 탐지를 위한 향후 목표를 설명한다.

2. 관련 연구

2.1 가짜뉴스 탐지

현재 국내에서는 다양한 웹 사이트에서 팩트 체크 서비스를 제공하고 있다. 대표적인 SNU FactCheck[7]는 서울대학교 언론정보연구소가 운영하고 있으며, 협업한 언론사들이 검증한 내용을 사람들에게 알리기 위한 목적을 가진다. SNU FactCheck에서의 검증 결과는 '전혀 사실 아님', '대체로 사실 아님', '절반의 사실', '대체로 사실', '사실', '판단 유보'와 같이 총 6가지로 나뉜다.

이와 더불어 국내외를 막론하고 인공지능 기반의 가짜뉴스 탐지 연구가 활발하게 진행되고 있다. 가짜뉴스를 탐지하려는 목적은 같지만 탐지하려는 가짜뉴스의 정의에 따라 다양한 방안들이 제시되었다. 해외에서 제안한 가짜뉴스 탐지 방안으로는 '가짜뉴스의 문맥과 비슷하게 작성된 기사'를 가짜뉴스로 정의하고 기사의 내용을 CNN(Convolutional Neural Network) 모델에 학습시켜 가짜뉴스를 탐지하는 모델[8], '기사의 내용이 관련 기사들과 상반되는 기사'를 가짜뉴스로 정의하고 기사와 그에 맞는 검증기사들을 사용하여 가짜뉴스를 탐지하는 LSTM(Long Short-Term Memory) 기반 모델 DeClarE[3]가 있다. 제안 알고리즘의 가짜뉴스 탐지 모델과 유사한 방안으로 가짜뉴스를 탐지하는 DeClarE는 임베딩 모델 GloVe를 사용하여 기사와 검증기사를 임베딩하고, Dense Layer를 통한 벡터 간 Attention 메커니즘 계산으로 중요 부분을 알 수 있는 Attention Weights를 출력한다. Attention Weights는 LSTM을 통해 문맥을 파악하는 과정을 거친 검증기사 벡터와 내적 계산으로 합쳐져 분류 과정에 입력된다. 검증기사의 내용에서 기사와 관련 있는 부분을 강조하여 분류 라벨과 함께 모델을 학습시키는 방법이다.

국내의 경우 첫 번째로, 영어 데이터셋으로 실험하였으며, '기사 제목과 기사 본문의 내용이 연관성이 없는 기사'를 가짜뉴스로 정의하고 기사 제목과 기사 본문의 내용을 BERT (Bidirectional Encoder Representations from Transformers) 기반 모델을 통해 비교하여 가짜뉴스를 탐지하는 연구[9], 첫 번째와 같이 '기사 제목과 기사 본문의 내용이 연관성이 없는 기사'를 가짜뉴스로 정의하고 정의에 맞는 한국어 기사를 수집하여 GRU(Gated Recurrent Unit) 기반 모델을 통해 가짜뉴스를 탐지하는 연구[10] 등 가짜뉴스 탐지 기술적 접근이 이루어졌다.

2.2 KoBERT

BERT[11]는 인공지능 기반 자연어처리 분야에 등장한 Google의 강력한 언어모델이다. 대용량 Unlabeled data로 모델을 사전 학습하여 문맥을 파악할 수 있도록 구축된 모델이다. 사전 학습된 모델 BERT에 분류하고자 하는 task에 맞게 Labeled data로 추가 학습하여 분류 성능을 높인다.

BERT는 영어 데이터뿐만 아니라 다국어 대규모 데이터로 사전 학습된 모델 또한 제공받을 수 있다[12]. Wikipedia에서 많이 사용되는 언어 중 상위 104개 언어로 이루어진 데이터가 사용되었으며 해당 데이터에 한국어도 포함된다. 이러한 BERT의 다국어 모델은 국내 연구에서도 많이 사용되고 있다. 그러나 국내의 SKT T-Brain은 영어와 한국어 사이의

어휘 차이가 있음에도 기본적으로 영어를 학습하기 위해 개발됐던 BERT 모델로 한국어 문맥을 파악하는 것은 성능 한계가 있다고 판단했다. 이러한 성능 한계를 극복하기 위해 SKT T-Brain이 개발한 것이 KoBERT(Korean BERT)이다. KoBERT는 Wikipedia와 한국어 기사 등에서 수집한 수백만 개의 한국어 문장을 통해 사전 학습되었으며, 한국어의 불규칙한 언어 변화의 특성을 학습시키기 위해 데이터 기반 토큰화(Tokenization) 기법을 사용하여 기존 대비 27%의 토큰만으로 정확도 측면의 성능을 2.6% 이상 향상시켰다[13].

2.3 문서 요약

문서 요약 기술은 크게 추출 요약(Extractive Summarization)과 생성 요약(Abstractive Summarization)으로 나뉜다. 추출 요약은 문서 내에서 중요한 문장들을 선택하여 추출해내는 방법이며, 생성 요약은 인공지능을 기반으로 문서의 문맥을 보고 새로운 문장으로 요약해내는 방법이다[14].

진위여부를 판단해야 하는 기사 요약 시, 터무니없는 요약이 생성되면 기사와 검증기사를 비교하는 제안 알고리즘의 가짜뉴스 탐지 모델이 잘못 학습되기에 본 논문에서는 문서 내에서 핵심 문장을 추출하는 추출 요약을 사용한다.

3. 제안 알고리즘

3.1 연구 개요

본 논문의 제안 알고리즘은 요약된 기사와 검증기사를 비교하여 가짜뉴스를 탐지하는 방안으로, KoBERT 기반 가짜뉴스 탐지 모델을 사용한다. 본 논문의 실험은 Google Colab을 사용하였으며 실험을 위해 개발한 코드[15]는 Colab에 공개하였다. 본 논문의 제안 알고리즘 전체 구조는 Fig. 1과 같다. 제안 알고리즘의 가짜뉴스 탐지 모델 학습을 위해 SNU FactCheck에서 수집한 기사와 검증기사를 사용한다.

본 논문은 본문의 글자 수가 많은 기사들 중에서 가짜뉴스를 탐지하기 위해서는 전체 정보를 함축한 요약된 정보를 사용하는 것이 더욱 효과적일 것이라 판단했다. 제안 알고리즘의 가짜뉴스 탐지에 사용하는 KoBERT는 입력 데이터의 길이가 토큰 512개로 제한되기 때문에, 길이가 긴 기사의 경우 최대 길이를 초과한 부분을 학습할 수 없다. 제안 알고리즘은 기사와 검증기사를 함께 입력하는 방안이기에 KoBERT에 입력될 두 기사는 각각 더 짧은 길이로 요구된다. 본 논문에서는 입력 데이터 길이 제한을 해결하기 위해 최대 길이에 맞게 강제로 기사의 내용을 삭제하지 않고, 요약을 통해 기사에서

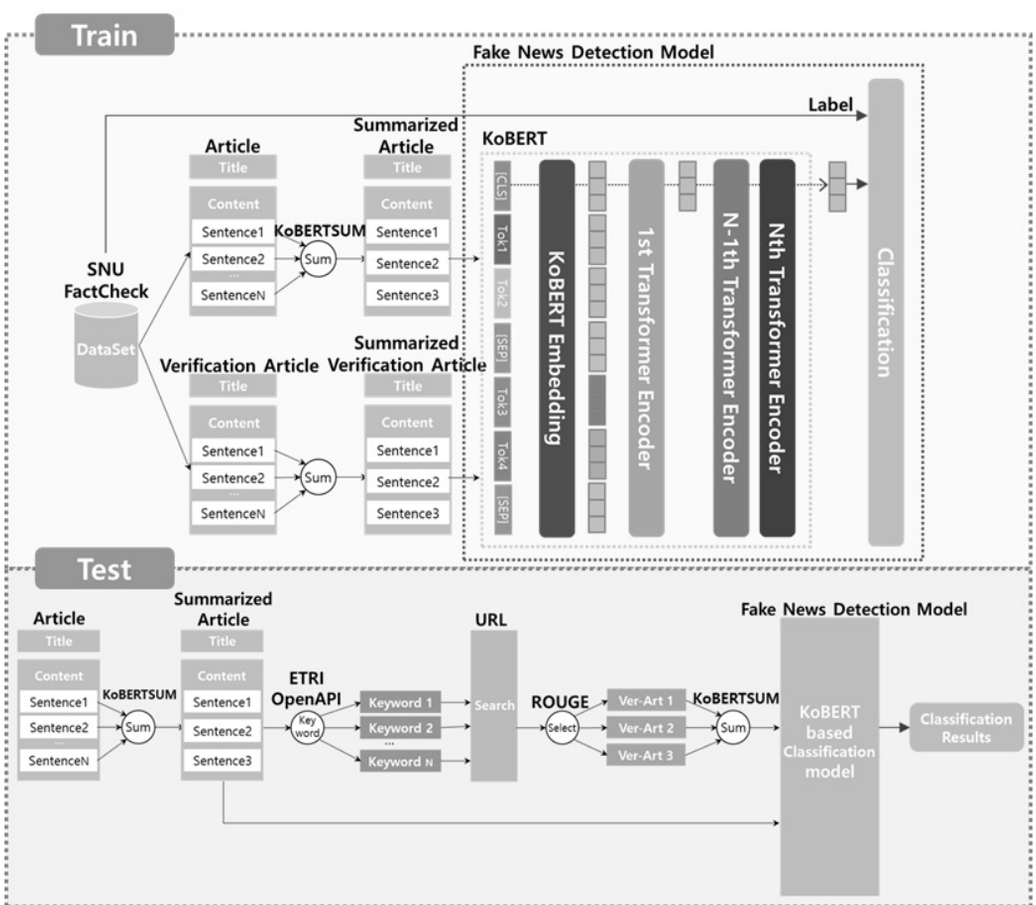

Fig. 1. Overview of the Algorithm Structure

소수의 문장을 추출하여 사용한다.

따라서 본 논문에서는 정확도 측면에서 성능이 뛰어나다고 알려져 있는 추출 요약 모델 BERTSum[16]을 통해 문서 요약 과정을 진행한다. BERTSum은 BERT를 통해 문서의 문맥을 파악하여 문장의 벡터를 생성한 후, Transformer에서 문장 벡터들의 비교를 통해 중요하다고 판단되는 문장을 분류하여 총 3문장의 추출 요약문을 생성하는 문서 요약 모델이다. 본 논문에서는 제안 알고리즘의 문서 요약 단계에서 BERTSum이 한국어 문서 요약 또한 가능하도록 BERT를 KoBERT로 변경하여 사용한다.

Fig. 2는 제안 알고리즘의 의사코드이다. 요약된 기사와 검증기사는 기사의 진위여부인 라벨과 함께 KoBERT 기반의 가짜뉴스 탐지 모델에 입력되며, 입력된 두 기사의 관계성 계산을 거쳐 생성된 최종 벡터는 라벨과 함께 KoBERT 위에 추가된 분류 모델 학습에 사용된다.

제안 알고리즘 테스트 단계에선 먼저 판단 대상인 기사를 KoBERTSum을 통해 요약한 뒤, ETRI의 언어 분석 API[17]를 통해 주요 키워드를 추출한다. 언어 분석 API는 문장에 쓰인 형태소 분석을 통해 각 형태소마다 사용 횟수 순으로 키워드를 추출한다. 본 논문에서는 명사 위주로 키워드를 추출하여 사용했다. 추출된 키워드는 검증기사 검색 단계에 사용되어 다양한 URL에서 관련 기사를 검색한다. 본 논문에서는 검증기사 검색을 위해 파이썬 오픈소스 라이브러리 BeautifulSoup[18]와 requests[19]를 사용했다. HTTP 요청을 통해 URL의 HTML 내용을 응답받는다. HTML 내용에서 원하는 부분을 추출하여 수집한다. 본 논문에서는 검증기사를 검색하는 각 URL마다 기사의 제목과 내용을 담은 태그에 맞게 검증기사 검색 코드를 작성하였다.

```
input: Articles α, Verification articles β, Classification answer γ

Initialize KoBERT model
Initialize KoBERTSum model

sum_α ← Summary of Articles using KoBERTSum;
sum_β ← Summary of Verification articles using KoBERTSum;

%TRAINING
for Each training epoch do
  PREDICT ← KoBERT(Tokenizer(sum_α, sum_β))
  Convert Tokens into word embedding vectors;
  Compute a representation of the sequence;
  Learn the classification model;

  if PREDICT == γ
    Then pass
  else
    perform training procedure to update weights;
End

%TEST
δ ← Search for verification articles;
sum_δ ← Summary of Search articles using KoBERTSum;
PREDICT ← KoBERT(Tokenizer(sum_α, sum_δ))
```

Fig. 2. Algorithm Pseudocode

검색된 검증기사들과 판단 대상인 기사 사이의 문서 품질 평가 척도 ROUGE 점수를 계산한 뒤, 결과를 기반으로 정확도가 높은 3개의 기사를 검증기사로 선택한다. 각 검증기사 또한 KoBERTSum을 통해 요약되며, 요약한 기사와 요약한 검증기사들을 학습된 가짜뉴스 탐지 모델의 입력 데이터로 기사의 진위여부를 판단한다.

3.2 데이터셋

본 논문의 실험을 위해서 기사와 해당 기사의 진위여부를 검증할 수 있는 검증기사가 필요하다. 진위여부가 확실하게 판단된 기사를 사용하여 학습하기 위해 언론사들이 수동으로 검증을 거친 SNU FactCheck의 기사를 수집하였다.

제안 알고리즘은 요약 데이터로 가짜뉴스 탐지 모델을 학습하기에 기사 내용이 하나의 주제를 다루고 있는 형태인 기사들을 위주로 수집했다. 인터뷰 기사의 경우 문서 요약 모델의 성능이 뛰어나다고 하더라도 잘못된 발언이 담긴 부분으로 요약되지 않으면 학습에 문제가 생길 수 있으므로 제외하였다. SNU FactCheck의 데이터가 많지 않고 주제 집중 기사만 사용할 경우 그 수가 더 적어 커뮤니티 게시물과 언론사 문제제기에 대한 데이터도 수집하였다. 온라인 기사는 아니지만, 언론사 측에서 팩트 체크할 대상 설명을 위해 온라인 기사 형태로 직접 작성한 내용이 있어 이를 활용하였다.

수집 결과, 가짜뉴스 데이터 651개, 진짜뉴스 데이터 294개로 총 945개의 데이터를 수집했다. '전혀 사실 아님'과 '사실' 데이터만 활용하기엔 데이터 수가 많지 않기에 가짜뉴스 데이터로는 '전혀 사실 아님' 데이터 507개에 '대체로 사실 아님' 데이터 144개를 추가로 수집했으며, 진짜뉴스 데이터로는 '사실' 207개에 '대체로 사실' 87개를 추가로 수집했다. 총 945개의 데이터들은 ['기사 제목', '기사 본문', '검증기사 제목', '검증기사 본문'. '분류 라벨']로 이루어져 있다. 수집한 데이터는 Table 1과 같이 학습 및 테스트 데이터로 분류하여 사용하였다.

4. 실 험

4.1 문서 요약 기법 검증

본 논문에선 가짜뉴스 탐지 모델에 입력되기 전 모든 기사를 요약한다. 사용한 문서 요약 모델인 KoBERTSum의 기반

Table 1. Dataset of Fake News Detection

Data	Label	Count
Train(90%)	Fake news	593
	Real news	257
Test(10%)	Fake News	58
	Real news	37

이 되는 BERTSum은 추출 요약 모델로써 좋은 성능을 보인다고 알려져 있지만, 진위여부를 판단하려는 기사를 요약할 때 기사의 주요 부분으로도 문제없이 요약이 되는지 검증했다. 문서 요약 모델 학습에는 DACON의 한국어 문서 추출요약 AI 경진대회[20]에서 제공된 약 4만 개의 기사 요약 데이터를 사용하였다.

문서 요약 모델 검증을 위해 모두 가짜뉴스로 이루어진 50개의 데이터를 사용했다. 데이터는 SNU FactCheck의 기사와 검증기사를 기반으로 생성했으며 [‘기사 본문’, ‘기사 정답 요약문’, ‘검증기사 본문’, ‘검증기사 정답 요약문’]으로 이루어져 있다. 기사 정답 요약문과 검증기사 정답 요약문은 본 논문에서 언론사에서 제공한 요약문 및 판단 주제를 참고하여 수작업으로 생성한 정답 요약문이다. 기사 본문과 검증기사 본문은 평균 약 10개의 문장으로, 기사 정답 요약문과 검증기사 정답 요약문은 3개의 문장으로 이루어져 있다. 문서 요약 모델을 통해 요약된 기사는 데이터의 기사 정답 요약문과 비교하고, 요약된 검증기사는 검증기사 정답 요약문과 비교했다.

Table 2는 기사 모델 요약문과 기사 정답 요약문 사이, 검증 기사 모델 요약문과 검증기사 정답 요약문 사이의 ROUGE 문서 품질 평가[21] 점수를 계산한 결과이다. 품질 평가에는 Rouge-1, Rouge-2, 그리고 Rouge-L 총 세 가지의 계산이 이루어졌다. ROUGE는 각 계산의 기준에 맞게 겹치는 단어의 수를 기반으로 품질 평가를 한다. 해당 실험에선 세 가지의 결과 중 겹치는 두 단어의 수를 계산하는 Rouge-2가 가장 낮게 측정되었다.

Table 3은 기사 모델 요약문과 검증기사 모델 요약문 사이의 내용을 비교한 결과이다. 기사에는 판단 대상이 되는 내용이 포함되어 있는지, 검증기사에는 기사의 진위여부를 검증할 수 있는 내용이 포함되어 있는지 검증했다.

Table 2. Average Rouge Score

Type	Rouge-1	Rouge-2	Rouge-L
Article	0.663	0.562	0.645
Verification article	0.647	0.542	0.628

Table 3. Number of Key Sentences

Type	Count	Number of Key Sentences		
		1-sen	2-sen	3-sen
Article Summary	Total Data (50)	12	21	17
	Rouge-2 0.5 or less (21)	9	10	2
Verification Article Summary	Total Data (50)	8	15	27
	Rouge-2 0.5 or less (24)	6	7	11

Article	Verification Article	Article Summary	Verification Article Summary
북한의 평창 겨울 올림픽 참 이런 가운데 미국 당국에선 유엔 대북 제재 결의 원칙이 현재 한국 정부는 우리가 초 당장 최문순 강원도지사는 국제올림픽위원회(IOC)도 " 2014년 인천아시안게임 때 이 때문에 통일부가 남북협 하지만 미국은 북한 선수단 북한의 핵·미사일 도발에 유엔 안보리의 대북 제재 결 우리 정부가 북한 선수단이 이에 미 당국은 한국이 정부 다만 미국 정부는 평창올림	평창올림픽에 참가하는 북한 미국무부도 유엔 대북제재이 남북 합의문 가운데에는 올 2016년 통과한 유엔 대북 제 다시 말해, 단돈 1원이라도 다만, 북측에 현금은 전달하 최문순 강원지사가 북한 방· 유엔 결의는 북한에 선박을 크루즈선을 보낼 수 있다는 과거 부산아시안게임 때는 ' 고려항공은 화물 검색을 받 북한에 들른 선박은 일정 기 그래서 논란을 피하려면 하·	북한의 평창 겨울 올림픽 참 이런 가운데 미국 당국에선 유엔 대북 제재 결의 원칙이	평창올림픽에 참가하는 북한 2016년 통과한 유엔 대북 제 미국무부도 유엔 대북제재이

Fig 3. Example of Summary Data

검증 결과, ROUGE 문서 품질 평가 결과와 상관없이 모든 기사에 판단 대상과 관련된 주요 내용이 포함됐다. 본 논문에서 가짜뉴스를 판단 대상이 되는 가짜뉴스 범위 특성 상, 주제 자체가 거짓인 기사로 이루어져있기 때문에 모델 요약문과 정답 요약문을 비교했을 때 다른 문장이지만 맥락이 일치하는 경우가 많았다. Table 3에서 나타난 내용처럼 모델 요약문에 주요 문장이 하나만 포함되는 데이터는 ROUGE 점수도 낮지만, 한 문장이 기사의 맥락을 표현하기엔 충분하다.

Fig. 3은 원문 데이터와 요약 데이터의 차이를 보인다. 해당 데이터는 '우리나라가 평창 올림픽에 참가하는 북한의 대표단, 태권도 시범단, 응원단 등에게 체재비를 지원하는 건 유엔 대북제재 결의 위반인가'에 대한 사실여부를 판단하기 위한 기사와 검증기사이다[22, 23]. 기사 요약문에는 ‘유엔 대북 제재 결의 원칙에 위반된다는 것이다.’와 같이 판단 대상 내용이 포함되었고, 검증기사 요약문에는 ‘2016년 통과한 유엔 대북 제재 결의 2270은 생계 같은 인도주의적 목적에 대해서만 대북 지원을 허락하는데 여기에도 조건이 붙기 때문에 군사적으로 전용 가능해서는 안 된다는 것.’과 같이 진위여부를 검증할 수 있는 내용이 포함되었기에 요약 데이터만으로도 진위여부를 판단할 수 있다.

본 논문에서는 요약문을 사용하여 가짜뉴스 탐지 모델을 학습하는 것의 효과를 검증하기 위하여 원문 데이터를 사용한 실험과 문서 요약 모델 KoBERTSum을 통해 요약한 데이터를 사용한 실험을 비교하였다. 실험에는 (1)BERT로 기사의 문맥만 파악하는 모델, (2)KoBERT로 기사의 문맥만 파악하는 모델, (3)BERT로 기사와 검증기사를 문맥을 비교하는 모델, (4)본 논문의 제안 알고리즘에서 사용하는 가짜뉴스 탐지 모델인, KoBERT로 기사와 검증기사의 문맥을 비교하는 모델 네 가지 모델을 사용했다.

Fig. 4는 네 가지 모델의 학습 과정에서 각 epoch의 손실값을 그래프로 나타낸 것이다. 손실값은 모델 학습 과정에서 손실 함수를 통해 계산된 모델의 예측값과 실제값의 차이, 즉 오차를 말한다. 손실값은 모델의 학습을 검증하기 위한 척도로 사용되어, 손실값을 줄이는 방향으로 학습을 진행한다. 모든 모델은 5번의 epoch로 학습했다.

네 가지 모델 모두 원문 데이터를 사용한 실험보다 요약 데이터를 사용했을 때 손실값이 더 빠르게 낮아졌다는 결과를 통해 요약 데이터가 모델 학습에 효율적임을 검증했다.

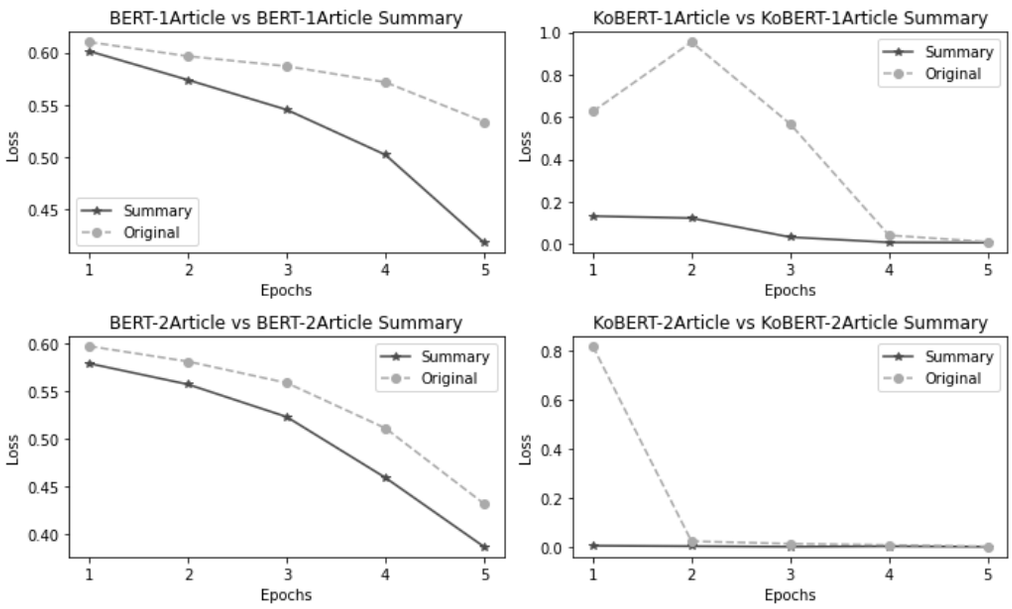

Fig. 4. Training Loss Curve of the Four Models

BERT와 KoBERT는 두 입력의 문맥을 비교하여 최종 벡터를 생성한다. 원문 데이터보다 요약 데이터를 사용했을 때 진짜뉴스와 가짜뉴스 사이의 최종 벡터 차이가 크게 생성되어 학습 과정에서 더욱 빨리 손실값을 줄일 수 있었던 것으로 보인다. 모든 실험에 학습률과 배치 크기는 동일하게 설정하였다.

본 논문에서는 요약의 필요성 증명을 위해, 테스트 과정에서 사용되는 기사들의 문장 개수에 따라 데이터를 분류하여 결과를 비교하였다. Table 1에서 설명한 테스트 데이터는 총 95개이며, 이를 기사의 문장 개수로 정렬하여 '길이가 긴 기사' 48개와 '길이가 짧은 기사' 47개로 분류하였다. '길이가 긴 기사'들은 평균 11개의 문장으로, '길이가 짧은 기사'들은 평균 7개의 문장으로 이루어졌다. Table 4는 본 논문의 제안 알고리즘의 가짜뉴스 탐지 모델에 원문 데이터와 요약 데이터를 사용하여 각각 학습시켜 두 개의 학습 모델을 생성한 후 실험을 진행한 결과이다.

실험 결과, 원문 데이터로 학습한 모델보다 요약 데이터로 학습한 모델의 예측 정확도가 높았다. 학습 데이터의 요약 유무는 모델의 학습에 영향을 주었다. 또한 두 개의 모델 모두 '길이가 긴 기사'보다 '길이가 짧은 기사'를 사용한 테스트 실험에서의 정확도가 높았다. 기사의 길이는 가짜뉴스 탐지에 영향을 미치며, 가짜뉴스 탐지 정확도 개선을 위해 요약 과정이 필요하다.

Table 4. Accuracy Comparison of Fake News Detection by Article Length

Test Article Length	Training Data	
	Unsummarized data	Summarized data
Long	72.91%	75.00%
Short	77.08%	79.16%

4.2 검증기사 검색 기법 검증

본 논문에서 제안 알고리즘 최종 검증에 사용할 검증기사는 연합뉴스, 중앙일보, 한국일보 등 SNU FactCheck 검증기사들의 출처인 다양한 URL에서 검색한 검증기사를 사용한다. 검증기사 검색에는 웹 크롤링 기술을 사용하였다. 해당 실험은 검증기사 검색 기법을 통해 가져온 검증기사들이 검증기사의 역할을 할 수 있는지 검증하기 위해 진행하였다. 기사에서 ETRI의 언어 분석 API를 통해 추출한 기사의 키워드를 사용하여 수행한 검증기사 검색에서 각 기사마다 검색된 3개의 검색기사와 SNU FactCheck에서 해당 기사의 진위여부를 판단하는데 사용된 검증기사의 내용을 비교하여 검색된 기사 또한 검증에 사용되기 적합한지 검증했다.

먼저 검색기사 검색 기법 환경을 구축하였다. 본 논문에서는 SNU FactCheck에서 사용하는 검증기사들의 출처 중 많은 부분을 차지하는 8개의 URL을 선택하였다. 여러 개의 기사를 사용한 검증기사 검색 테스트를 통해 각각의 URL 특성에 맞도록 코드를 작성하였다.

생성한 검증기사 검색 기법 환경을 검증하기 위하여 환경 구축에 사용된 데이터에 포함되지 않은 30개의 데이터를 사용한 실험을 진행했다. 실험 데이터는 SNU FactCheck의 ['기사 제목+기사 본문', '검증기사 제목+검증기사 본문']으로 이루어져있다. 진짜뉴스 데이터 15개와 가짜뉴스 데이터 15개를 사용했으며, Fig. 5는 라벨마다 기사 2개의 검증기사 검색 결과만을 표시한 내용이다. 각 기사를 통해 검색된 3개의 검증기사들과 기존의 검증기사의 내용을 비교했을 때 검

Article	Verification article	Search article	Label	Suitable for Use
		홍준표 "국민의힘, 민	Real	O
홍준표 "국민의힘, 민	홍준표 "국민의힘은	홍준표 "당 추구하는	Real	O
		홍준표, 국민의힘에 "	Real	O
		김신조 사건' 52년 만	Real	O
문 대통령, 52년 만에	52년 만에 개방' 북	김신조 사건' 52년 만	Real	O
		문 대통령, 52년 만에	Real	O
		공무원은 5인 이상 사	Fake	O
공무원 5인 이상 사적	공무원은 5인 이상 사	"공무원에게만 5인 이	Fake	O
		코로나 확산세 꺾였다	Fake	O
		日 역사 교과서, '독도	Fake	O
조법종 "일본 역사 교	日 역사 교과서, '독도	日 모든 교과서 '독도	Fake	O
		日 중학 全 사회교과	Fake	O

Fig. 5. Web Crawling Results

색기사가 검증에 사용될 수 있는지에 대한 여부를 나타낸다.

사용 가능한 검색기사의 수가 1개인 데이터는 9개, 2개인 데이터는 10개, 3개인 데이터는 11개 존재했다. 본 논문에서는 검증기사 검색 기법 검증 실험을 통해 모든 기사에서 1개 이상의 검증기사를 검색했다.

4.3 가짜뉴스 탐지 모델 정확도 검증

본 논문에서 제안하는 알고리즘의 가짜뉴스 탐지 모델 성능을 검증하기 위하여 유사한 방안의 모델과 비교 실험을 진행하였다. 비교에 사용된 모델은 DeClarE이다.

DeClarE는 기사와 검증기사 데이터만 사용하는 본 논문과 다르게 출처 데이터도 함께 사용하는 모델이지만, 동일한 데이터셋으로 비교하기 위해 임베딩된 출처 벡터를 추가하는 과정은 생략하고 진행하였다. 또한 DeClarE의 임베딩에 사용된 GloVe의 경우, 영어로만 사전 학습되어 있는 모델이기 때문에 본 논문에서는 GloVe에 한국어 기사를 학습시켜 한국어 데이터도 임베딩 할 수 있도록 생성하였다.

또한 제안 알고리즘에서 사용하는 가짜뉴스 탐지 모델의 기반이 되는 KoBERT의 성능을 검증하기 위하여 KoBERT의 비교 대상인 BERT의 다국어 모델을 사용한 가짜뉴스를 탐지 실험도 진행했다. 실험에는 DeClarE, BERT, 그리고 KoBERT까지 총 세 가지의 모델을, 학습 및 테스트에는 모두 Table 1에서 설명한 데이터셋을 사용했다.

세 가지 모델에 각 두 가지 실험을 진행했다. 첫 번째 실험은 요약을 하지 않은 기존의 데이터셋을 그대로 사용하였고, 두 번째 실험은 데이터 전체에 KoBERTSum을 사용하여 요약을 한 데이터셋을 사용하였다. 실험 결과, 모든 모델에서 요약하지 않은 데이터를 사용한 실험보다 요약된 데이터를 사용한 실험에서 가짜뉴스 탐지 정확도가 높았다.

Table 5. Accuracy Comparison of Fake News Detection Model

Model	Unsummarized data	Summarized data
DeClare	58.52%	63.85%
BERT	72.63%	73.68%
KoBERT	75.00%	77.08%

모델 간의 예측 정확도를 비교해본 결과, DeClarE와 BERT 모델보다 KoBERT의 정확도가 높았다. DeClarE의 임베딩에 사용되는 GloVe는 저차원 벡터에 단어의 의미를 표현할 수 있지만 단어 수준 임베딩이기 때문에 문장 전체의 맥락을 파악하여 임베딩하는 것이 어려운 반면, BERT와 KoBERT의 경우 전체 문맥을 고려하여 임베딩하기에 입력 데이터 임베딩 과정부터 학습에 영향을 준 것이다.

또한 DeClarE의 하나의 레이어를 통한 Attention 계산과 다르게 BERT와 KoBERT의 경우, 네트워크의 Multi-Head Attention[24] 단계를 통해 여러 개 Head가 각각의 관점으로 Attention 계산을 진행하여 특정한 하나의 위치에만 크게 집중하는 것이 아니라, 그 집중을 분산시킬 수 있어 다양한 위치를 효과적으로 학습한다. 이와 같은 이유로 DeClarE보다 BERT와 KoBERT가 더 높은 정확도를 보인 것으로 판단되며, BERT보다 KoBERT의 정확도가 높은 이유는 BERT의 다국어 모델보다 KoBERT가 한국어 학습에 수월하도록 사전 학습되었기 때문이라고 판단된다.

4.4 제안 알고리즘 최종 검증

해당 절에서는 학습된 제안 알고리즘에 새로운 기사를 사용하여 제안 알고리즘을 검증한다. 제안 알고리즘이 자동화 시스템으로 활용되기 위해 기사와 검증기사 요약, 올바른 키워드 검출, 검증기사의 검색이 제안 알고리즘의 성능에 미치는 영향, 제안한 가짜뉴스 탐지 모델의 예측 성능을 검증한다. 본 검증은 기사 요약을 시작으로 검증기사 검색 기법을 통한 검증기사 선택과, 가짜뉴스 탐지 모델의 예측까지 모든 과정을 진행했다. 검증에는 SNU FactCheck의 진짜뉴스 15개와 가짜뉴스 15개 총 30개의 기사를 사용하였으며 모든 기사는 10개의 문장으로 이루어져 있다.

Table 6은 실험 결과이다. 라벨마다 5개 기사 결과를 나타냈다. 기사 및 검증기사 요약 단계에서 요약 결과에 판단 대상과 관련된 주요 내용이 1문장 이상 포함되었기에 결과 내용을 생략했다. Table 6의 속성 중 'person'은 기사와 관련된 특정 인물이 존재하는지를 나타낸 것이다. 검증기사 검

Table 6. Result of Experiment Using SNU FactCheck Data

Label	Art	Subject	person	Verification Articles Count	Successful detection
Real	1	COVID	O	2	X
	2	Politics	O	3	O
	3	International	O	3	X
	4	Politics	X	1	X
	5	Politics	O	2	O
Fake	6	Economy	X	3	O
	7	Society	X	2	X
	8	Politics	O	2	X
					O
	9	Society	X	1	O
	10	COVID	O	3	O

Table 7. Result of Experiment Using Another Data

Art	Subject	Verification Articles Count	Successful detection
1	COVID	3	O
2	Society	2	O
3	Society	0	-
4	Economy	3	O
5	COVID	1	O
6	COVID	3	O
7	Politics	3	O
			X
			O
8	Politics	1	X
9	Economy	0	-
10	Politics	2	O

색 결과, 모든 기사에서 1-3개의 사용 가능한 검증기사가 선택되었다. 사용 가능한 검증기사가 2개 이상 선택된 기사들의 특징을 분석해보면, 검증기사가 다수 선택될 수 있었던 이유는 코로나 및 유명 이슈를 주제로 다뤄 관련 기사들을 많이 찾아볼 수 있거나, 기사와 관련된 특정 인물이 존재하여 검증기사 검색 단계에서 관련 기사 검색이 수월했기 때문이다. 검증에 사용된 30개의 기사 중 코로나 관련 기사 6개는 검증기사 검색을 통해 모두 2개 이상의 검증기사를 선택하였고, 기사 내용에 특정 인물의 이름이 들어가는 기사 12개도 모두 2개 이상의 검증기사를 선택하였다.

선택된 검증기사를 사용하여 가짜뉴스 탐지 모델로 예측한 결과, 총 65개의 예측 중 41개가 예측에 성공했다. 검증기사가 올바르게 선택되었으나 탐지에 실패한 경우, 제안한 가짜뉴스 탐지 모델은 정치 카테고리로 분류된 데이터가 90% 이상 존재하는 수집 데이터로 학습된 모델이기에 국제, 사회 등 다른 카테고리에서 사용된 문맥을 파악하기에 어려움이 있었다. Fig. 6은 검증에 사용된 기사들의 주제와 각 주제마다의 예측 성공 횟수 그래프다. 정치 관련 기사와 달리 다른 주제

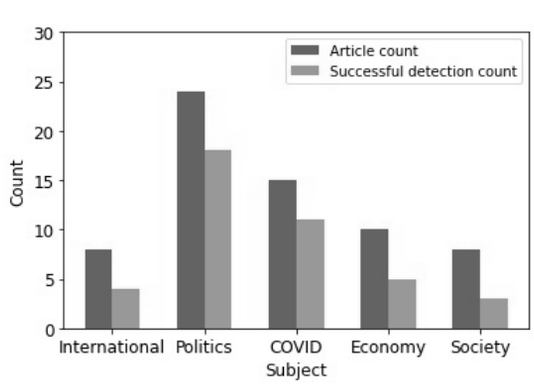

Fig. 6. Subject of Articles in SNU FactCheck Data

기사의 가짜뉴스 탐지 예측은 성공 확률이 낮았다.

같은 기사에서 검색된 검증기사들은 가짜뉴스 탐지 과정에서 대부분 서로 동일한 예측 결과가 나왔으나, Table 6에서 기사 8번의 검증기사 2개는 서로 다른 예측 결과를 보였다. 검증기사를 분석한 결과, 예측에 실패한 검증기사의 경우 내용에 판단 대상 기사를 검증하는 내용을 담고 있지만, 제목은 판단 대상 기사의 제목과 거의 유사한 문장으로 작성되어 진실로 예측한 것으로 판단된다.

본 논문에선 SNU FactCheck에서 검증이 완료된 기사뿐만 아니라, 다른 팩트 체크 사이트인 팩트체크넷[25]에서 검증을 진행한 내용에 대해서도 제안 알고리즘으로 가짜뉴스 탐지가 가능한지 실험을 통해 검증했다.

팩트체크넷은 SNU FactCheck와 다르게 언론사뿐만 아니라 시민 팩트체커로 선정된 시민이 함께 협업하여 판단 대상을 제시하고, 검증 자료를 제공한다. 실험에는 팩트체크넷에서 시민 팩트체커가 진행한 팩트 체크 내용 30개를 사용하였다. 시민 팩트체커가 제시한 판단 대상은 2-3 문장으로 짧게 요약되어 있었으므로 제안 알고리즘의 요약 과정을 생략하고 검증기사를 검색하기 위한 키워드 추출 단계부터 진행하였다.

Table 7은 실험 결과이다. 라벨마다 5개 기사 결과를 나타냈다. 사용 가능한 검증기사의 수는 SNU FactCheck의 판단 대상을 사용한 실험의 사용 가능한 검증기사의 수보다 적게 나타났다. 제안 알고리즘은 SNU FactCheck의 판단 대상들을 기반으로 검증기사가 쉽게 검색될 수 있도록 환경을 구축하였으나, 팩트체크넷의 시민 팩트체커들은 언론사 링크를 제공하지 않고 검증 대상을 작성한 경우가 많기 때문에 이에 대한 검색이 어려웠다.

본 논문에서는 검증기사 검색 단계에서 검증기사가 하나도 검색되지 않은 판단 대상 기사 3개의 내용을 분석했다. 팩트

체크넷은 해당 내용을 시민으로부터 제안 받아 시민 팩트체커가 검증을 진행하였다. 검증을 위해 전문기관 자료 및 공공데이터 분석 결과를 제시하였으며 본 논문에서 수작업으로 관련 기사를 찾아본 결과, 해당 내용을 다룬 기사가 없었다.

검색된 검증기사들을 사용하여 가짜뉴스 탐지 모델을 통해 가짜뉴스 분류를 한 결과, 58개의 예측 중 41개가 예측에 성공했다. 실험 결과, 제안 알고리즘은 판단 대상에 따라 온라인 기사만으로 검증하기 어려운 내용을 제외하면 SNU FactCheck의 데이터 외에도 가짜뉴스 탐지가 가능하였다.

5. 결 론

본 논문은 기사 데이터만 사용하는 경우, 가짜뉴스 조작의 정밀도가 높아졌을 때 탐지가 어려워질 수 있다는 한계를 극복하기 위해 기사와 함께 사용될 검증기사를 확보하여 활용하는 가짜뉴스 탐지 방법을 제안하였다. 제안 알고리즘은 기사와 검증기사 데이터가 사용되기 전에 요약이 되도록 요약 과정을 추가하였으며, 실험을 통해 요약 데이터 사용 시 가짜뉴스 탐지 모델의 정확도가 향상됨을 증명하였다. 또한 제안 알고리즘의 가짜뉴스 탐지 모델과 유사한 방안의 모델과의 비교 실험을 통해 제안 알고리즘의 성능을 증명하였다. 본 논문은 가짜뉴스 조작의 발달에 영향을 받지 않기 위해 새로운 추가 데이터를 사용했다는 점에서 의의가 있다고 판단하였다.

제안 알고리즘은 향후 자동화를 위해 검증기사 검색을 통해 선택한 기사 중 가짜뉴스가 포함되지 않도록 검증기사 검색 기법의 보완이 필요하다. 향후, 가짜뉴스의 출처와 날짜 정보를 추가로 활용하여, 가짜뉴스가 많았던 URL은 검증기사 선택 확률이 낮아지도록 개선할 것이다. 또한, 검증기사로 온라인 기사뿐만 아니라 소셜 미디어 및 법률, 공공데이터 등에서 검증 내용을 가져와 검증할 수 있는 범위를 넓히고자 한다.

References

[1] H. Jwa, D. Oh, and H. Lim, "Research analysis in automatic fake news detection," *Journal of the Korea Convergence Society*, Vol.10. No.7, pp.15-21, 2019.

[2] Y. Yoon, T. Eom, and J. Ahn, "Fake news detection technology trends and implications," Weekly Technology Trends, Institute of Information & Communications Technology Planning & Evaluation, Oct. 2017.

[3] K. Popat, S. Mukherjee, A. Yates, and G. Weikum, "DeClarE: Debunking fake news and false claims using evidence-aware deep learning," *EMNLP*, pp.22-32, 2018.

[4] H. Lee, J. Kim, and J. Paik, "Survey of fake news detection techniques and solutions," *Proceedings of the Korean Society of Computer Information Conference*, pp.37-39, 2020.

[5] Y. Hyun and N. Kim, "Text mining-based fake news detection using news and social media data," *The Journal of Society for e-Business Studies*, Vol.23, No.4, pp.19-39, 2018.

[6] J. Shim, J. Lee, I. Jeong, and H. Ahn, "A study on Korean fake news detection model using word embedding," *Proceedings of the Korean Society of Computer Information Conference*, Vol.28, No.2, pp.199-202, 2020.

[7] SNU FactCheck [Internet], https://factcheck.snu.ac.kr/.

[8] R. Kumar, A. Goswami, P. Narang, and S. Sinha, "FNDNet – A deep convolutional neural network for fake news detection," *Cognitive Systems Research*, Vol.61, pp.32-44, 2020.

[9] J. Heejung and O. Dongsuk, "exBAKE: Automatic fake news detection model based on bidirectional encoder representations from transformers (BERT)," *Applied Sciences*, Vol.9, No.4062, 2019.

[10] S. Yoon, K. Park, and J. Shin, "Detecting incongruity between news headline and body text via a deep hierarchical encoder," *arXiv:1811.07066*, 2018.

[11] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *arXiv:1810.04805*, 2018.

[12] BERT [Internet], https://huggingface.co/transformers/model_doc/bert.html.

[13] KoBERT [Internet], https://github.com/SKTBrain/KoBERT.

[14] Text Summarization with Attention mechanism [Internet], https://wikidocs.net/72820.

[15] Source Link [Internet], https://drive.google.com/drive/folders/1kQmSoYoq8AoXsbmstTpL2zLjPUEHQB1o?usp=sharing.

[16] Y. Liu, "Fine-tune BERT for Extractive Summarization," arXiv:1903.10318v2, 2019.

[17] ETRI Language Analysis API [Internet], https://aiopen.etri.re.kr/guide_wiseNLU.php.

[18] BeautifulSoup [Internet], https://www.crummy.com/software/BeautifulSoup/bs4/doc/.

[19] Requests [Internet], https://pypi.org/project/requests/.

[20] Korean document extractive summarization AI competition [Internet], https://dacon.io/competitions/official/235671/overview/description.

[21] C. Lin, "ROUGE: A Package for Automatic Evaluation of Summaries," Text Summarization Branches Out, pp.74-81, 2004.

[22] The JoongAng News Article [Internet], https://www.joongang.co.kr/article/22260785#home.

[23] SNU FactCheck ‘"Verification article," [Internet], https://factcheck.snu.ac.kr/v2/facts/435.

[24] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, and L. Jones, "Attention Is All You Need," *arXiv:1706.03762*, 2017.

[25] FactChectNet [Internet], https://factchecker.or.kr/.

한 윤 진

https://orcid.org/0000-0002-1675-7193
e-mail : qkr030@naver.com
2013년 동의대학교 영상정보공학과(학사)
2019년 ~ 현 재 동의대학교 디지털미디어공학과 석사과정
관심분야 : 탈 중앙 웹, 인공지능, 설명가능 인공지능, 블록체인, 자기주권 데이터

김 근 형

https://orcid.org/0000-0002-7691-5608
e-mail : geunkim@deu.ac.kr
1986년 서강대학교 전자공학과(학사)
1988년 서강대학교 전자공학과(공학석사)
2005년 포항공과대학교 컴퓨터공학과 (공학박사)
2007년 ~ 현 재 동의대학교 게임공학전공 교수
관심분야 : 탈 중앙 웹, 인공지능, 설명가능 인공지능, 블록체인, 자기주권 데이터