

뉴스와 소셜 데이터를 활용한 텍스트 기반 가짜 뉴스 탐지 방법론

Text Mining-based Fake News Detection Using News And Social Media Data

현윤진(Yoonjin Hyun)*, 김남규(Namgyu Kim)**

초 록

최근 가짜 뉴스가 분야를 막론하고 전 세계에서 주목을 받고 있으며, 현대경제연구원에서는 이러한 가짜 뉴스로 인한 피해 규모가 연간 약 30조 900억원에 달하는 것으로 추산하였다. 정부에서는 “가짜 뉴스 찾기”를 주제로 “인공지능 R&D 챌린지” 대회를 개최하여 가짜 뉴스를 가려낼 인공지능 원천기술 개발에 대한 첫 걸음을 내딛고 있으며, 민간 차원에서도 다양한 분야에서 팩트 체크 서비스가 제공되고 있다. 학계에서도 가짜 뉴스를 탐지하기 위한 시도가 전문가 기반, 집단지성 기반, 인공지능 기반, 시맨틱 기반 등으로 활발하게 이루어지고 있다. 하지만 이러한 시도는 조작의 정밀도가 높을수록 뉴스 자체에 대한 분석만으로 진위 여부를 식별하기가 더욱 어렵다는 한계를 경험하고 있으며, 가짜 뉴스 탐지 모델의 정확도가 과평가된 경향을 보이고 있다. 따라서 본 연구에서는 가짜 뉴스 탐지 모델 정확도의 공정성을 확보하고, 뉴스의 내용뿐만 아니라 해당 뉴스에 대한 반응으로 자연적으로 발생한 광범위한 소셜 데이터를 활용하여 뉴스의 진위 여부를 판정하는 방안을 제안하고자 한다.

ABSTRACT

Recently, fake news has attracted worldwide attentions regardless of the fields. The Hyundai Research Institute estimated that the amount of fake news damage reached about 30.9 trillion won per year. The government is making efforts to develop artificial intelligence source technology to detect fake news such as holding “artificial intelligence R&D challenge” competition on the title of “searching for fake news.” Fact checking services are also being provided in various private sector fields. Nevertheless, in academic fields, there are also many attempts have been conducted in detecting the fake news. Typically, there are different attempts in detecting fake news such as expert-based, collective intelligence-based, artificial intelligence-based, and semantic-based. However, the more accurate the fake news manipulation is, the more difficult it is to identify the authenticity of the news by analyzing the news itself. Furthermore, the accuracy of most fake news detection models tends to be overestimated. Therefore, in this study, we first propose a method to secure the fairness of false news detection model accuracy. Secondly, we propose a method to identify the authenticity of the news using the social data broadly generated by the reaction to the news as well as the contents of the news.

키워드 : 텍스트 분석, 토픽 모델링, 가짜 뉴스 탐지

Text Analysis, Topic Modeling, Fake News Detection

이 논문은 2017년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2017S1A5A2A03067632).

* First Author, Graduate School of Business IT, Kookmin University(yoonjin0630@kookmin.ac.kr)

** Corresponding Author, School of Management Information Systems, Kookmin University
(ngkim@kookmin.ac.kr)

Received: 2018-10-22, Review completed: 2018-11-08, Accepted: 2018-11-18

1. 서 론

최근 “가짜 뉴스(Fake News)”가 분야를 막론하고 전 세계에서 주목을 받고 있다. 2016년 옥스퍼드 사전이 “탈 진실(Post-Truth)”을 올해의 단어로 선정한 것에 이어, 2017년에는 영국 사전출판사 콜린스가 전년도 대비 사용빈도가 365% 증가한 “가짜 뉴스”를 올해의 단어로 선정하는 등 국내외적으로 가짜 뉴스에 대한 관심이 급증하고 있다. BuzzFeed에 의하면, 미 대선 관련 상위 20개 가짜 뉴스 중 17개가 조작되었으며, 가짜 뉴스에 대한 페이스북 반응률이 주류 언론기관의 진짜 뉴스에 대한 페이스북 반응을 상회하는 것으로 나타났다. 이러한 가짜 뉴스는 정치적·경제적 목적을 가지고 양산되고 있으며, 에이 따라 개인을 비롯한 기업과 국가 역시 엄청난 사회적 비용을 치루고 있다.

최근 현대경제연구원에서는 이러한 가짜 뉴스로 인한 개인, 기업, 그리고 사회적 피해 금액이 연간 약 30조 900억 원에 달하는 것으로 추산한 바 있다. 한편 이러한 보고서의 내용에 대해서도 “가짜 뉴스 피해 30조 원이라는 가짜 뉴스”라는 언론의 반박이 있을 정도로, 가짜 뉴스를 둘러싼 사회적 담론은 매우 혼란스러운 양상으로 전개되고 있다. 가짜 뉴스가 확산하는 원인 중 하나는 생산 비용이 비교적 저렴하다는 점으로, MIT 테크놀로지 리뷰에 따르면 약 20만 달러의 비용으로 길거리 시위를 선동할 수 있다고 한다. 또한 텍스트 조작뿐 아니라 음성, 동영상 조작까지 가능해지는 등 가짜 뉴스 생성을 위한 기술이 더욱 정교화, 고도화됨에 따라, 일반적인 방법으로는 원본과 가짜의 구분이 불가능한 시물라시옹 시대가 이미 도래하였다고 해도 과언이 아니다.

이러한 배경에서 가짜 뉴스에 대한 사회적 문제를 진단하고 이로 인한 부작용을 막기 위한 시도들이 다양한 분야에 걸쳐 활발하게 이루어지고 있다. 기존의 전문가 기반 탐지, 집단 지성 기반 탐지 등 비기술적 접근뿐 아니라, 인공지능 기반 탐지, 시맨틱 기반 탐지 등 기술적 접근도 활발하게 이루어지고 있다. 비기술적 접근의 경우, 검증 과정과 결과가 명확하고 공신력이 있다는 장점이 있으나, 소수 전문가의 지식에 전적으로 의존함으로써 정보에 대한 전수조사가 어렵고 상대적으로 많은 분석 시간이 요구된다는 한계를 갖는다. 집단지성 기반 탐지의 경우에도 사용자들의 적극적인 참여가 전제되어야 한다는 측면, 그리고 검증 결과의 신뢰성을 보장할 수 없다는 측면의 한계를 갖는다. 한편 기술적 접근의 경우 이러한 비기술적 접근의 한계를 보완할 수 있지만, 방대한 양의 학습 데이터가 요구되거나 지식 데이터 축적을 위해 상당한 비용과 노력이 요구된다는 또 다른 한계를 갖는다.

또한 이들 연구는 조작의 정밀도가 높을수록 뉴스 자체에 대한 분석만으로 진위 여부를 식별하기가 더욱 어려워진다는 한계를 공통적으로 경험하고 있다. 실제로 최근 수행되고 있는 가짜 뉴스 탐지 연구의 대부분은, 진정한 의미의 가짜 뉴스가 아닌 광고성 뉴스, 제목과 기사의 내용이 다른 뉴스 등 일종의 스팸 뉴스를 탐지하는 것에 초점을 두고 있다. 스팸 뉴스의 경우, 어휘 패턴 분석, 제목과 내용의 유사도 분석 등을 통해 비교적 쉽게 식별이 가능하지만, 가짜 뉴스의 경우 그 내용의 진위를 뉴스 자체만으로 판정하려는 시도는 본질적인 한계를 갖는다.

이처럼 가짜 뉴스 탐지 연구 분야는 여러 한계를 가지고 있는 것은 물론, 정확도 검증의 신뢰성 역시 보장할 수 없다. 예를 들어 <Figure 1>은 모두 2014년에 발생했던 “세월호 침몰” 관련 뉴스이다. A, B, C 뉴스 모두 비슷한 이슈를 가지고 있으나, A, B 뉴스의 경우, 표현만 다를 뿐 결국 “승객 구조”에 대한 같은 내용을 전달하고 있다. 반면 C 뉴스의 경우, A, B 뉴스와 비슷한 이슈이나 “수색작업 종료”에 대한 내용을 전달하고 있다. 일반적으로 A 뉴스가 가짜 뉴스로 판명됐을 때, 이를 토대로 B와 C 뉴스를 예측하려는 시도가 기존 연구들의 흐름이다. 하지만 B 뉴스의 경우, A 뉴스와 원문만 다를 뿐 동일한 내용을 다루고 있기 때문에 진위여부 역시 동일하다. 반면 C 뉴스의 경우, A 뉴스와 비슷한 이슈를 가지고 있기는 하나 내용이 다르기 때문에 해당 뉴스의 진위여부를 판단하는 것은 어렵다. 이러한 측면에서 본다면 가짜 뉴스 탐지 모델의 정확도를 평가함에 있어, A 뉴스를 학습 데이터로 하여 C 뉴스를 예측하는 것은 문제가 없지만, A 뉴스를 학습 데이터로 하여 B 뉴스를 예측하는 것은 정확도가 과평가 될 가능성이

높다. 또한 특정 이슈가 지속적으로 이어지게 되면 매우 유사한 이슈가 동일한 진위여부를 가질 수밖에 없고, 이러한 이슈를 갖는 뉴스들이 학습 및 검증 데이터로 사용되는 경우에도 정확도가 과평가 될 수 있기 때문에 시기를 나누어 예측하는 것도 바람직하지 않다.

실제 이러한 현상이 발생하는지 확인해보기 위해 본 연구에서 실제 뉴스 데이터를 대상으로 파일럿 실험을 수행한 결과가 <Figure 2>에 나타나있다. <Figure 2>는 뉴스 데이터의 학습 및 검증 데이터 집합 구분 기준에 따른 가짜 뉴스 예측 정확도로, 학습과 검증 데이터를 무작위로 나누어 예측을 한 경우, 90% 이상의 정확도를 보이고 있음을 확인할 수 있다. 또한 동일한 데이터를 대상으로 시기를 구분하여 예측을 한 경우에도 약 80%의 정확도를 보이고 있음을 확인할 수 있다. 이는 뉴스 기사들의 이슈들 사이에 단단하게 얽혀있는 연결 관계(Tightly Connection)로 인한 정확도의 과평가 가능성을 반증한다. 따라서 가짜 뉴스 예측 모델의 엄밀한 검증을 위해서는 이슈 간 연결 관계를 끊어내고, 보다 엄격한 환경을 조성(Harsh Setting)할 필



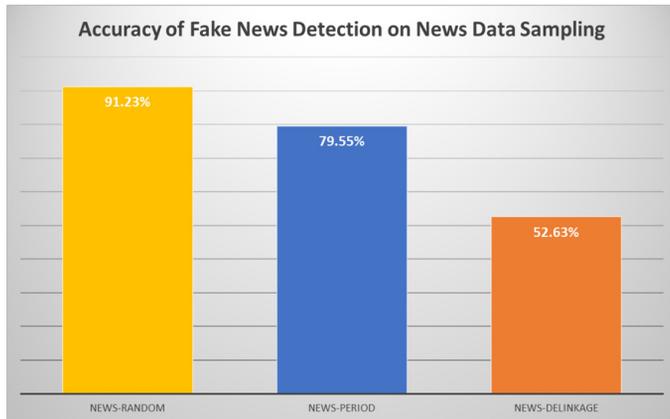
<Figure 1> Identification of News Authenticity

요가 있다. <Figure 2>에서 실제 뉴스 데이터 간 이슈 연결 관계를 차단하여 예측한 결과, 정확도가 약 52% 정도로 떨어진 것을 확인하였다.

하지만 뉴스 기사들의 이슈 간 연결 관계를 끊어내고 뉴스 자체만을 대상으로 진위여부를 학습하여 새로운 뉴스의 진위여부를 예측하는 것은 매우 어려운 일이다. 따라서 본 연구에서는 뉴스의 내용은 물론, 해당 뉴스에 대한 반응으로 자연적으로 발생한 광범위한 소셜 데이터를 활용하여 뉴스의 진위 여부를 판정하는 방

안을 제안하고자 한다.

<Figure 3>은 2017년 5월 2일 보도된 “세월호 인양 고의지연 의혹”에 대한 내용으로, 보도 직후 가짜 뉴스임이 밝혀져 언론사 측에서 사과 방송을 내보냈던 가짜 뉴스의 대표적 사례이다. 한편 그림 좌측은 해당 뉴스 보도 직후부터 가짜 뉴스로 판명되기 전까지 실제 발생한 Tweet으로, “장난치지마”, “뉴스 마디마디가 왜곡되어 연결”, “악질 가짜 기사”, “거짓뉴스 대응” 등 보도 내용에 대해 다양한 반응을 보임



<Figure 2> Accuracy of Fake News Detection on News Data Sampling



<Figure 3> Examples of Fake News Detection Using Social Media Data

을 확인할 수 있다. 이는 실시간으로 확산되는 소셜미디어 상의 여론으로부터 뉴스의 진위여부에 대한 판단의 단초를 찾을 수 있음을 암시한다. 즉, 기존의 뉴스 자체만으로는 식별하기 어려운 뉴스의 진위여부를 소셜미디어 상의 여론을 추가로 활용함으로써 좀 더 정확하게 진위여부를 식별할 수 있을 것으로 판단하였다. 이에 본 연구에서는 1. 뉴스 기사 사이의 이슈간 연결 관계를 끊어낸 엄격한 환경에서 정확도를 측정하고, 2. 뉴스 내용과 더불어 Twitter로부터 각종 뉴스에 대한 반응들을 활용하여 사실이 아닌 허위 정보를 다루는 가짜 뉴스를 더욱 정확하게 탐지하기 위한 방안을 제시한다. 구체적으로는 각 뉴스와 관련된 Tweet을 추려내고, 이들로부터 각 뉴스에 대한 반응들을 벡터 형태로 산출한 후, 뉴스 원문에서 추출한 기사의 주제적 특성(Topic Features)과 트위터 반응 벡터(Twitter Topic Feature)와의 결합을 통해 각 뉴스의 진실/거짓 여부를 더욱 정확하게 탐지하고자 한다.

2. 관련 연구

2.1 토픽 모델링

토픽 모델링(Topic Modeling)은 텍스트 마이닝의 응용 분야 중 하나로써 다양한 분야에서 가시적인 성과를 내며 가장 활발하게 활용되고 있는 대표적인 응용 기술이다[15]. 토픽 모델링은 각 문서에 포함된 용어의 빈도수에 기반하여 유사 문서를 그룹화한 뒤 각 그룹을 대표하는 주요 용어들을 추출함으로써 해당 그룹의 토픽 키워드 집합을 제시하는 방식으로 이루어지며,

이 때 사용되는 문서는 문서, 제목, 요약, 본문, 댓글 등을 포함하는 넓은 개념을 의미한다. 주요 이론적 배경으로는 벡터공간모델(Vector Space Model)[1, 17]과 TF-IDF(Term Frequency-Inverse Document Frequency)[19]이 있다. TF-IDF는 여러 문서에서 자주 출현하는 일반적인 단어에 대한 가치는 낮게 부여하고 특정 문서에서만 출현하는 특수한 단어에 대해 가치를 높게 부여하는 방식으로 이루어지며, 각 문서는 용어 수만큼의 차원을 갖게 되며 TF-IDF를 값으로 갖는 벡터로 표현된다. 이때, TF-IDF 값은 다음과 같은 방법으로 산출되며, 다양한 분야에서 활용되어지고 있다[16].

$$TF-IDF(d, t) = TF(d, t) \times IDF(t)$$

$$TF(d, t) = \begin{cases} 0 & \text{if } freq(d, t) = 0 \\ 1 + \log(1 + \log(freq(d, t))) & \text{otherwise} \end{cases}$$

$$IDF(t) = \log \frac{|d|}{|d_t|}$$

2.2 가짜 뉴스 탐지

가짜 뉴스와 그에 따른 피해 사례가 급증함에 따라 국내외를 막론하고 가짜 뉴스를 탐지하기 위한 시도가 활발하게 수행되고 있다. 국회입법조사처의 “이슈와 논점”에서는 2016년 미 대선의 가짜 뉴스 논란을 집중 조명하였으며, 이후 인터넷 괴담 및 가짜 뉴스를 막기 위한 괴담방지법·가짜 뉴스 방지법 등에 대한 국회의원들의 입법 발의가 이루어지고 있다. 지난해 있었던 19대 대선에서는 선거관리위원회에서 가짜 뉴스를 단속하기 위한 “비방·흑색선전 전담 TF팀”을 구성하여 운영하였으며, 과학기술정보통신부에서는 “가짜 뉴스 찾기”를 주제로 “인공지능 R&D 챌린지” 대회를 개최하여

진짜 뉴스와 가짜 뉴스를 가려낼 인공지능 원천기술 개발에 대한 첫 걸음을 내딛고 있다.

한편, 민간 차원에서도 국내외 다양한 분야에서 팩트 체크(Fact Check) 서비스가 제공되고 있다. 현재 국내에서는 안철수 팩트 체크 센터, 아리송닷컴, 뉴스톱, SNU FactCheck 등 10여 개의 팩트 체크 서비스가 운영되고 있다. 특히 SNU FactCheck는 언론사와 대학이 협업하는 공공 정보 서비스 모델로서, 서울대학교 언론정보연구소에서 웹플랫폼을 마련하고 해당 서비스에 참여하는 언론사들이 팩트 체크가 필요한 정보에 대해 교차 검증을 하는 방식으로 운영되고 있다. 해외의 경우 2018년 3월 현재 위키피디아 기준으로 전 세계 38개국에서 PolitiFact, FactCheck.org 등 24개의 팩트 체크 서비스가 제공되고 있다. 또한 가짜 뉴스 탐지를 위한 민간의 노력으로 Facebook의 경우 Facebook 내의 가짜 뉴스 문제를 해결하기 위해 실시간으로 해당 정보의 출처를 확인하는 인공지능 기반의 가짜 뉴스 탐지 도구를 선보인 바 있으며, Google의 경우, Crosscheck 프로젝트를 발표하여 프랑스 주요 언론사들의 참여를 통해 뉴스의 진위를 판단하고 이를 검색 알고리즘에 적용하였다.

이와 더불어 학계에서도 가짜 뉴스를 탐지하기 위한 연구가 활발하게 수행되고 있다. 해외의 경우, 나이브 베이지안 분류기를 적용한 가짜 뉴스 탐지 연구[5], 언어 정보 및 SVM을 네트워크 기반의 행동 데이터와 결합한 방식의 연구[4], 뉴스 기사의 텍스트 및 클릭 배이트(Click Bait) 단서를 결합한 방식의 가짜 뉴스 탐지 연구[3], 이상 확산 패턴 감지와 시맨틱 분석을 결합한 하이브리드 방식의 가짜 뉴스 탐지 연구[11] 등 기술적 접근이 활발히 이루어지고 있다

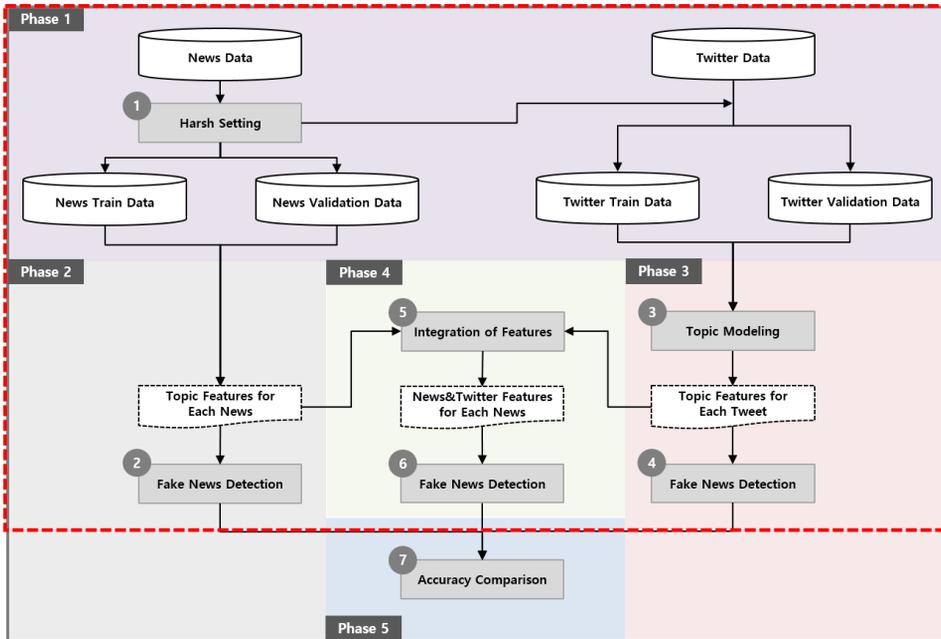
[2, 7, 14, 18]. 한편 국내의 경우, 주로 가짜 뉴스의 개념 정립, 사례 및 현황 분석, 규제 방안 등에 대한 비기술적 접근의 연구가 주를 이루고 있다[10, 12, 6]. 하지만 최근 들어 토픽 모델링을 통한 가짜 뉴스의 탐색적 연구[9], 가짜 뉴스의 포털과 SNS 간 반응 차이 분석에 대한 연구[8], SVM을 이용한 가짜 뉴스 판별 시스템에 대한 기술적 접근[13] 등이 이루어지고 있다.

3. 가짜 뉴스 탐지 방법론

3.1 연구 개요

본 절에서는 가짜 뉴스 탐지 방법론을 제안한다. <Figure 4>는 본 연구의 전체 개요도를 나타내고 있으며, 원통형으로 표시된 부분은 데이터 소스를 나타낸다. 또한 직사각형으로 표시된 부분은 주요 프로세스를 나타내며, 검정색 점선으로 표시된 도형은 각 프로세스의 산출물을 나타낸다.

제안 방법론은 총 4부분으로 구성되어 있으며(Phase 1~Phase 4), Phase 1과 Phase 4가 본 연구의 핵심이라고 할 수 있다. Phase 1은 가짜 뉴스 탐지 모델 정확도의 과평가 방지를 위해 뉴스 기사의 이슈 간 연결 관계를 차단하여 뉴스 데이터를 분할하는 부분으로, 뉴스 데이터를 대상으로 토픽 모델링을 통해 구조화한 결과를 기반으로 클러스터링을 수행함으로써 이슈 그룹을 생성하여 동일 이슈이나 진위여부가 다른 뉴스 기사들을 학습과 검증 데이터로 분리시키는 방식으로 이루어진다. 이렇게 분리된 뉴스 기사에 따라 관련 트위터 데이터들 역시 해당 뉴스 기사의 진위여부에 종속되어 학습과 검증



〈Figure 4〉 Research Overview

데이터로 분리되어 가짜 뉴스 탐지 모델의 대상 데이터로 활용된다. 이 과정을 통해 가짜 뉴스 탐지 모델 구축을 위한 엄격한 환경이 제공되며, 이후 뉴스와 트위터 데이터를 활용한 가짜 뉴스 탐지 모델을 구축하게 된다. Phase 2는 뉴스 데이터만을 활용한 가짜 뉴스 탐지 모델을 구축하는 부분으로, Phase 1의 결과로 분리된 뉴스 데이터의 주제적 특성을 활용하여 가짜 뉴스 예측 모델을 구축한다. Phase 3은 트위터만을 활용한 가짜 뉴스 탐지 모델을 구축하는 부분으로, Phase 2와 마찬가지로 Phase 1의 결과로 분리된 트위터 데이터를 대상으로 토픽모델링을 수행하여 트위터 반응에 대한 벡터 값을 추출하고, 뉴스별 차원별 평균 벡터 값을 산출해 활용함으로써 가짜 뉴스 예측 모델을 구축한다. 마지막 Phase 4는 본 연구의 핵심 부분 중 하나로 뉴스와 트위터 데이터를 활용한 가짜 뉴스 탐지 모델을

구축하는 부분이다. 구체적으로는 Phase 2와 Phase 3에서 추출된 뉴스에 대한 주제적 특성과 트위터 반응에 대한 벡터 값을 결합하여 가짜 뉴스 탐지 모델을 구축한다. 이후 제 4.3절에서 Phase 5의 정확도 비교를 통하여 본 제안 방법론에 대한 예측 정확도를 평가하게 된다. 제안 방법론에 대한 보다 자세한 설명은 이어지는 제3.2절~제3.4절에서 다루도록 한다.

3.2 Phase 1: 정확도 과평가 방지를 위한 데이터 분할

본 절에서는 제안 방법론의 핵심 중 하나인 가짜 뉴스 탐지 모델 정확도의 과평가 방지를 위한 데이터 분할 부분으로(Phase 1), 뉴스 기사의 이슈 간 연결 관계 차단을 통한 엄격한 실험 환경 조성에 대해 소개한다. 구체적으로

는 뉴스 데이터를 대상으로 토픽모델링을 통해 뉴스를 구조화하고, 그 결과를 클러스터링하여 이슈 그룹을 생성함으로써 학습과 검증 데이터를 분리하는 과정을 설명한다.

<Figure 5>는 뉴스 기사의 이슈 간 연결 관계를 차단하여 학습 및 검증 데이터를 분리시키는 과정을 보여주는 가상의 예이다. <Figure 5>의 (a) 뉴스 데이터를 대상으로 토픽모델링을 통해 구조화하면 (b)와 같은 문서/토픽 대응 행렬을 얻을 수 있다. 이 때, 각 행렬의 셀 값은 특정 토픽에 대한 문서의 부합 정도를 나타내는 문서 가중치(Document Weight)를 나타내며 Phase 2의 입력 값으로 활용될 뉴스의 주제적 특성을 의미한다. 이후, 이 결과를 활용하여 클러스터링을 수행함으로써 유사한 이슈들을

갖는 뉴스 그룹을 생성하게 된다. 이때, 생성된 각 이슈 그룹들 내 뉴스 기사들의 진위여부에 따라 학습(Train)과 검증(Validate) 데이터로 분리하게 되며 <Figure 5>의 (c)를 보면 뉴스 기사 2번과 4번이 “세월호” 관련 이슈를 가지고 동일한 이슈 그룹에 속해있음에도 진위여부가 “Fake”와 “True”로 각각 다르기 때문에 학습과 검증 데이터로 분리된 것을 알 수 있다.

이후, 분리된 뉴스 기사에 따라 트위터 데이터 역시 학습과 검증 데이터로 분리된다. 이때, 트위터 데이터는 뉴스 기사와 종속 관계를 가지고 있기 때문에 관련된 뉴스의 진위여부에 따라 해당 트윗(Tweet)의 진위여부가 결정된다. 예를 들어, <Figure 5>의 (a) 뉴스 기사 중 2번 뉴스 기사에 관련된 트위터가 10개라고 가

Target	newsID	Title
Fake	1	Rescue all of Danwon highschool student in Sewolho
Fake	2	Sewolho salvage delay suspicion
True	3	Sewolho 3 cycle memorial ceremony
True	4	Sewolho salvage completed
Fake	5	Parliamentary agreement needed to declare the end of inter-Korean war
True	6	North and South Summit Panmunjom Declaration
Fake	7	Chu Ja-Hyun unconscious state
True	8	Chu Ja-Hyun discharge
Fake	9	Hong Joon-pyo remarks "Roh Moo Hyun Bibbe Account"
True	10	Hong Joon-pyo resigns from the party
...

(a) News Data

Target	newsID	N_Topic1	N_Topic2	N_Topic3	N_Topic4	N_Topic5
Fake	1	0.8	0.6	0	0	0
Fake	2	0.9	0.7	0	0	0
True	3	0.8	0.6	0	0	0
True	4	0.9	0.7	0	0	0
Fake	5	0	0	0.9	0	0
True	6	0	0	0.8	0	0
Fake	7	0	0	0	0	0.9
True	8	0	0	0	0	0.8
Fake	9	0	0	0	0.9	0
True	10	0	0	0	0.7	0
...

(b) News Data Structured Results

ClusterID	set	Target	newsID	N_Topic1	N_Topic2	N_Topic3	N_Topic4	N_Topic5
1	Train	Fake	1	0.8	0.6	0	0	0
	Train	Fake	2	0.9	0.7	0	0	0
	Validate	True	3	0.8	0.6	0	0	0
	Validate	True	4	0.9	0.7	0	0	0
2	Validate	Fake	5	0	0	0.9	0.4	0
	Train	True	6	0	0	0.8	0.3	0
3	Train	Fake	7	0	0	0	0	0.9
	Validate	True	8	0	0	0	0	0.8
4	Validate	Fake	9	0	0	0	0.9	0
	Train	True	10	0	0	0	0.7	0
...

(c) Separate Training and Validation Data Through Issue Group Creation

<Figure 5> Example of Harsh Setting

<Table 3> Virtual Prediction Results of A Fake News Detection Model Using News and Twitter Data

newsID	N_Topic1	...	N_Topic5	T_Topic1	...	T_Topic5	Target	Prediction
3	0.8	...	0	0.23	...	0	True	Fake
4	0.9	...	0	0	...	0	True	True
5	0	...	0	0.1	...	0	Fake	Fake
8	0	...	0.8	0	...	0.72	True	Fake
13	0.3	...	0	0	...	0	True	True
15	0.2	...	0	0.2	...	0	Fake	Fake
20	0	...	0	0.2	...	0	True	True
21	0	...	0	0.5	...	0	Fake	Fake
...

가상의 예측 결과를 보여주고 있다. <Table 2>를 보면 뉴스 기사 4번이 <Table 1>에서와 마찬가지로 “Fake”로 잘못 예측된 것을 확인할 수 있으며, 이는 트위터 데이터만으로 뉴스의 진위여부를 판정하는데 한계가 있음을 보여준다.

3.5 Phase 4: 뉴스와 트위터 데이터를 활용한 가짜 뉴스 탐지 모델

본 절에서는 본 연구의 핵심인 Phase 4의 뉴스와 트위터 데이터를 활용한 가짜 뉴스 탐지 모델에 대해 소개하며, 구체적으로는 Phase 1과 Phase 3을 통해 산출된 뉴스별 뉴스의 주제적 특성과 트위터 벡터 값을 활용하여 기계학습 알고리즘을 사용해 가짜 뉴스 탐지 모델을 구축하는 과정을 설명한다.

뉴스와 트위터 데이터를 모두 활용한 가짜 뉴스 탐지 모델을 구축하기 위해서는 뉴스 데이터의 속성과 트위터 데이터의 속성을 결합하는 과정이 우선적으로 수행되어야 한다. 이때, 뉴스 데이터의 속성은 Phase 1의 과정을 통해 뉴스별 주제적 특성이 산출된 것을 활용하며, 트위터 속성의 경우에는 Phase 3을 통해 산출

된 뉴스별 트위터 벡터 값을 활용한다. 뉴스의 주제적 특성과 뉴스별 트위터 벡터 값을 결합하는 과정은 뉴스의 주제적 특성에 뉴스별 트위터 벡터 값들을 새로운 변수로 추가하는 방식으로 이루어진다.

이후, 뉴스의 주제적 특성과 트위터 벡터 값이 결합된 산출물을 입력 값으로 하여 기계학습 알고리즘을 사용해 가짜 뉴스 탐지 모델을 구축하게 된다. <Table 3>은 가상의 예측 결과를 보여주고 있으며, Phase 2와 Phase 3의 예측 모델에서는 모두 “Fake”로 잘못 예측했던 뉴스 기사 4번이 뉴스와 트위터 데이터를 모두 활용한 가짜 뉴스 탐지 모델로 예측했을 때는 “True”로 제대로 된 예측을 한 것을 확인할 수 있다.

4. 실험

4.1 실험 데이터 소개

본 연구의 실험을 위해서는 진위여부가 식별된 뉴스 데이터와 해당 뉴스와 관련된 트위터 데이터가 필요하다. 그런데 진위여부가 식별된

데이터의 경우, 진위여부가 식별된 뉴스 데이터 수 자체가 매우 적을 뿐만 아니라, 진위여부가 식별된 뉴스의 경우에도 가짜 뉴스의 비율이 상대적으로 높다. 또한 가짜 뉴스라고 식별된 뉴스의 경우, 실제 뉴스 데이터 보다는 정치인 혹은 공직자 등의 발언, 찌라시, 루머 등에 대한 내용이 주를 이루고 있어 원문 내용이 어려운 것은 물론, 실제 뉴스 데이터임에도 불구하고 가짜 뉴스로 판명된 이후에는 대부분 원문 내용 혹은 관련 내용이 삭제된 경우가 대부분이어서 데이터 수집에 많은 어려움이 따른다. 더불어 뉴스의 진위여부가 식별된 시기가 모호하기 때문에 관련 트위터 데이터를 수집함에 있어 기간에 대한 기준을 세우기도 애매한 것이 사실이다. 따라서 본 연구에서는 가짜 뉴스를 의도와 상관없이 허위 사실을 포함하고 있거나 허위 사실을 전달하는 실제 뉴스 데이터 혹은 허위 사실을 말하고 있는 정치인 혹은 공직자의 발언, 찌라시, 루머 등을 포함하는 뉴스 데이터로 정의하여 데이터를 수집하였다. 또한 뉴스 기사의 진위여부 식별 기준을 팩트체크 사이트 중 기준사이트를 선정하여 해당 사이트에서 판단 치우침이 “대체로 거짓”, “거짓”인 경우를 “거짓”, “대체로 진실”, “진실”인 경우를 “진실”이라고 정의하였다. 더불어 원문 혹은 관련 내용이 나타나있지 않은 경우, 네이버 뉴스에서 키워드 검색을 통해 가장 유사하고 대표적인 뉴스 데이터를 수집하였고, 상대적으로 부족한 진실 뉴스 데이터의 경우에는 실제로 일어난 사건에 대한 뉴스 데이터를 수집하였다.

뉴스 데이터의 경우, 가장 활발히 운영되고 있으며 공신력이 있다고 판단되는 SNU FactCheck 과 뉴스톱 2개의 사이트를 기준사이트로 선정하

였고, SNU FactCheck에서는 2017년 3월 29일~2018년 6월 22일 사이의 총 679건 중 가짜 19건, 진실 9건을 수집하였으며, 뉴스톱에서는 2017년 4월 27일~2018년 6월 22일 사이의 총 333건 중 가짜 36건, 진짜 9건을 수집하였다. 부족한 데이터는 2017년 1월 1일~2018년 6월 22일 사이의 뉴스들 중 실제로 일어난 사건에 대한 뉴스 혹은 명백히 가짜로 밝혀진 뉴스 데이터 총 61건(가짜 1건, 진짜 60)을 수집하여, 총 134건(가짜 56건, 진실 78건)의 뉴스 데이터를 수집하였다.

트위터 데이터의 경우, 앞서 수집된 뉴스 데이터에 대해 트위터의 고급검색 기능을 활용하여 키워드 검색을 통해 관련 트윗 총 16,384건을 수집하였다. 이때, 뉴스 진위여부 식별의 시기가 모호하기 때문에 수집 기간을 일괄적으로 2016년 1월 1일~2018년 7월 10일로 한정하여 적용함으로써 데이터를 수집해 실험에 사용하였다.

4.2 실험 결과

4.2.1 Phase 1: 정확도 과평가 방지를 위한 데이터 분할 결과

본 부절에서는 제3.2절에서 소개된 과정에 따라, 수집된 실제 뉴스 데이터와 트위터 데이터를 활용하여 뉴스 기사의 이슈 간 연결 관계를 차단함으로써 가짜 뉴스 탐지 모델 정확도의 과평가 방지를 위한 학습 및 검증 데이터 분리 과정 및 결과를 소개한다.

우선 전체 뉴스 데이터를 대상으로 토픽 모델링을 수행하여 뉴스 데이터를 구조화하였으며, 토픽의 수는 최대로 추출 가능한 26개로 지정하여 실험을 진행하였다. 이후, 토픽 모델링 결과로

산출된 문서/토픽 매트릭스, 즉 뉴스별 주제적 특성을 입력 값으로 하여 클러스터링을 수행함으로써 이슈 그룹 30개를 도출하였다. 이때, 이슈 그룹의 개수는 개수를 조정하며 실험을 한 결과, 이슈 그룹을 가장 잘 분리해 표현한 30개의 이슈 그룹을 실험에 사용하였다. 이렇게 생성된 이슈 그룹을 활용하여 각 이슈 그룹 내 뉴스 기사들의 진위여부에 따라 학습과 검증 데이터로 분리하는 작업을 수행하였고, 그 결과의 일부가 <Figure 7>에 나타나 있다.

<Figure 7>의 A와 B를 보면 같은 이슈 그룹 내에 있으면서도 서로 다른 진위여부를 갖기 때문에 각각 학습과 검증 데이터로 분리된 것을 확인할 수 있다. 또한 트위터 데이터 역시 뉴스 데이터의 학습 및 검증 구분에 따라 분리하여 저장하였다. 이에 따라 뉴스 데이터 총 134건 중 학습 데이터 77건(진실 45건, 거짓 32건), 검증 데이터 57건(진실 33건, 거짓 24건)으로 분리되었으며, 트위터 데이터 총 16,384건 중 학습 데이터 10,371건(진실 5,623건, 거짓 4,748

건), 검증 데이터 6,013건(진실 3,028건, 거짓 2,985건)으로 분리되어 가짜 뉴스 탐지 모델 구축에 활용되었다.

4.2.2 Phase 2: 뉴스 데이터를 활용한 가짜 뉴스 탐지 모델 구축 결과

본 부절에서는 제3.3절에서 소개된 과정에 따라, 제4.2.1절에서 산출된 뉴스 데이터의 학습 및 검증 데이터를 활용하여 가짜 뉴스 탐지 모델을 구축한 결과를 소개한다.

우선 제4.2.1절에서 분리된 뉴스 데이터 총 134건 중 학습 데이터 77건, 검증 데이터 57건을 대상으로 가짜 뉴스 탐지 모델을 구축하였다. 이때, 뉴스 데이터에 대한 구조화 작업이 이미 제4.2.1절에서 이루어졌으므로, 그 결과를 활용하여 가짜 뉴스 탐지 모델을 구축하였으며, 가짜 뉴스 탐지 모델을 구축하기 위한 기계학습 알고리즘은 파일럿 실험을 통해 가장 좋은 성능을 보였던 뉴럴 네트워크(Neural Network)를 사용하였다. <Table 4>는 예측 결과의 일부

ClusterID	set	Target	newsID	Title	Contents	Topic1	...	Topic26
9	Train	Fake	1	"문재인 부산에 800평 좋은 집"	박지원 "웃"	0.035	...	0.046
	Train	Fake	50	"문재인 정부가 지난 9년간 살린 살림을 들어먹고 있다"	홍준표 "9"	0.052	...	0.024
	Validate	True	66	댓글만으로 여론을 조작할 수 있는지?	드루킹이	0.053	...	0.047
13	Train	Fake	28	"자유한국당만 선거법위반 처벌"은 사실?	김성태 "합"	0.316	...	0.269
	Validate	True	53	교육감 후보는 정당과 기호가 없다?	중앙선관위	0.031	...	0.323
26	Train	True	11	문재인대통령 휴가	문재인 대	0.011	...	0.011
	Train	True	43	영화 규합도 스크린 독점했다?	규합도'로	0.014	...	0.027
	Train	True	77	추자현 퇴원	추자현 측	0.005	...	0.003
	Train	True	82	북미정상, 역사적 합의 서명	북미정상,	0.013	...	0.022
	Train	True	102	말레이-北 결국 무비자 협정 파기	말레이-北	0.004	...	0.005
	Validate	Fake	4	"음주운전 3회, 성희롱 2번이 정부 새 인사 기준"	음주운전	0.02	...	-0.004
	Validate	Fake	6	홍준표 "문준용 특혜 의혹 제기 안했다"	저는 지난	0.015	...	0.006
	Validate	Fake	12	대통령 연차휴가 21일 사용 위법"	문대통령	-0.002	...	0.003
	Validate	Fake	21	"미 선전포고로 전투기 격추는 자위권 행사"	北 리용호	0.008	...	0.033
	Validate	Fake	24	"좌파코드없는 영화는 천만 관객 불가능"	홍준표 자	0.035	...	0.053
Validate	Fake	54	"팀추월 당시 노선영이 마지막 주자 원했다"	백철기 감	0.02	...	0.009	
Validate	Fake	75	추자현, 출산후 열흘째 의식불명...산소호흡기 의지	추자현, 출	0.007	...	0.007	

<Figure 7> Results of Harsh Setting(Part)

(Table 4) Prediction Results of Fake News Detection Model Using News Data(Part)

newsID	Title	N_Topic1	...	N_Topic26	Target	Prediction
3	하태경 “문재인, 매춘부 합법화”	0.026	...	0.007	Fake	True
4	“음주운전 3회, 성희롱 2번이 정부 새 인사 기준”	0.021	...	0.013	Fake	True
6	홍준표 “문준용 특혜 의혹 제기 안했다”	0.014	...	0.002	Fake	True
12	대통령 연차휴가 21일 사용 위법	-0.002	...	0.000	Fake	True
21	“미 선전포고로 전투기 격추는 자위권 행사”	0.008	...	0.018	Fake	True
24	“좌파코드없는 영화는 천만관객 불가능”	0.035	...	0.002	Fake	Fake
53	교육감 후보는 정당과 기호가 없다?	0.033	...	0.020	True	Fake
54	“팁추월 당시 노선영이 마지막 주자 원했다”	0.020	...	0.020	Fake	True
66	댓글만으로 여론을 조작할 수 있는지?	0.050	...	0.072	True	Fake
75	추자현, 출산후 열흘째 의식불명...산소호흡기 의지	0.003	...	0.005	Fake	True
76	2018 남북정상회담 판문점 선언에 서명	0.017	...	0.017	True	Fake
100	‘뇌물 혐의’ 피의자로 전락한 현직 대통령	0.021	...	0.028	True	True
110	스웨덴서도 트럭 테러	-0.001	...	-0.080	True	True
113	거제 크레인 사고	0.001	...	0.039	True	True
114	의정부 상가서 큰 불	0.000	...	-0.005	True	True
126	냉장고 냉동실서 아기 시신 2구 발견	0.005	...	0.003	True	True
130	文대통령, SNS에 방미 첫날 후기 “한미동맹 더 강하게 발전할 것”	0.019	...	0.267	True	True
131	‘제보 조작’ 이유미 구속	0.024	...	0.022	True	True

를 보여주고 있으며, 제4.2.1절의 <Figure 7>에서 동일한 이슈를 가졌으나 진위여부가 달라 분리된 뉴스 학습 데이터 77번, 82번 뉴스 기사로 인해 뉴스 기사 21번, 75번이 “True”로 잘못 예측되었음을 확인할 수 있었다.

4.2.3 Phase 3: 트위터 데이터를 활용한 가짜 뉴스 탐지 모델 구축 결과

본 부절에서는 제3.4절에서 소개된 과정에 따라, 제4.2.1절의 결과로 분리된 트위터의 학습 및 검증 데이터를 활용한 가짜 뉴스 탐지 모델 구축 과정 및 결과를 소개한다.

우선 트위터 데이터 총 16,384건을 대상으로

토픽 모델링을 통해 구조화 작업을 수행하였으며, 토픽의 개수는 뉴스와 동일하게 26개로 지정하였다. 그 결과, 각 뉴스별 트윗에 대한 벡터 값이 산출되었고, 각 뉴스별 트윗의 차원별 평균값을 산출하여 최종적으로 뉴스별 트위터 벡터 값을 도출하였다. 이후, 제4.2.1절에서 분리된 학습 데이터 10,371건, 검증 데이터 6,013건으로 나누어 제4.2.2절과 마찬가지로 뉴럴 네트워크를 사용하여 가짜 뉴스 탐지 모델을 구축하였다. <Table 5>는 예측 결과의 일부를 보여주고 있으며, 제4.2.2절과 마찬가지로 21번, 75번 뉴스 기사가 여전히 “True”로 잘못 예측됨을 확인할 수 있었다.

(Table 5) Prediction Results of Fake News Detection Model Using Twitter Data(Part)

newsID	T_Topic1	...	T-Topic26	Target	Prediction
3	0.00637	...	0.00547	Fake	Fake
4	-0.01496	...	0.00541	Fake	True
6	0.00102	...	-0.00443	Fake	True
12	0.01140	...	0.00813	Fake	True
21	0.02103	...	0.00218	Fake	True
24	0.00115	...	0.00469	Fake	True
53	0.00045	...	0.00332	True	True
54	-0.00013	...	0.00090	Fake	True
66	-0.00010	...	0.00196	True	Fake
75	0.00032	...	0.00188	Fake	True
76	0.00205	...	0.00354	True	Fake
100	0.00982	...	-0.00427	True	Fake
110	-0.00134	...	0.00159	True	True
113	0.00000	...	0.00422	True	Fake
114	0.00000	...	0.00125	True	True
126	0.00016	...	0.00272	True	True
130	0.00451	...	0.00126	True	Fake
131	-0.00047	...	-0.00128	True	Fake

4.2.4 Phase 4: 뉴스와 트위터 데이터를 활용한 가짜 뉴스 탐지 모델 구축 결과

본 부절에서는 제3.5절에서 소개된 과정에 따라, 제4.2.2절에서 산출된 뉴스별 주제적 특성과 제4.2.3절에서 산출된 뉴스별 트위터 벡터 값을 결합하여 가짜 뉴스 탐지 모델을 구축한 과정 및 결과를 소개한다.

우선 뉴스별 주제적 특성에 뉴스별 트위터 벡터 값을 차원별로 새로운 변수로 추가하여 결합하였고, 이를 입력값으로 하여 뉴럴 네트워크를 사용해 가짜 뉴스 탐지 모델을 구축하였다. <Table 6>은 예측 결과의 일부를 나타내고 있으며, <Table 4>와 <Table 5>에서 모두 제대로 예측하지 못했던 뉴스 기사 21번과 75번이 “Fake”로 제대로 예측됨을 확인하였다.

4.2.5 Phase 5: 검증

본 제안 방법론(제4.2.4절)의 성능을 검증하기 위하여 제4.2.2절의 뉴스 데이터를 활용한 가짜 뉴스 탐지 모델과 제4.2.3절의 트위터 데이터를 활용한 가짜 뉴스 탐지 모델의 정확도를 비교 분석한 결과가 <Figure 8>에 나타나 있다. <Figure 8>에서 기존의 뉴스 데이터 자체만을 활용한 가짜 뉴스 탐지 모델과 트위터만을 활용한 가짜 뉴스 탐지 모델은 모두 50%대의 정확도를 보인 반면, 뉴스와 트위터 데이터를 결합한 가짜 뉴스 탐지 모델의 경우 80% 이상의 높은 정확도를 보이며 본 제안 방법론의 성능이 약 28%p 향상됨을 확인하였다.

본 제안 방법론의 성능이 향상됨을 보인 것은 매우 고무적인 현상이나, 그 폭이 너무 크기 때문에

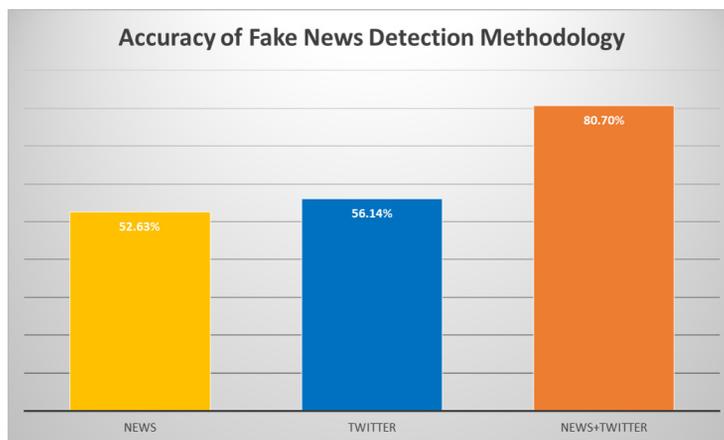
<Table 6> Prediction Results of Fake News Detection Model Using News and Twitter Data(Part)

newsID	N_topic1	...	N_topic26	T_topic1	...	T_topic26	Target	Prediction
3	0.026	...	0.007	0.006368421	...	0.005473684	Fake	True
4	0.021	...	0.013	-0.01496296	...	0.005407407	Fake	Fake
6	0.014	...	0.002	0.001016393	...	-0.00442623	Fake	Fake
12	-0.002	...	0	0.006397436	...	0.004551282	Fake	True
21	0.008	...	0.018	-0.00345724	...	0.008967105	Fake	Fake
24	0.035	...	0.002	0.002673469	...	-0.002244898	Fake	Fake
53	0.033	...	0.02	-0.00012821	...	0.000897436	True	Fake
54	0.02	...	0.02	0.015761905	...	0.007095238	Fake	Fake
66	0.05	...	0.072	-0.00067442	...	-0.000313953	True	True
75	0.003	...	0.005	0.002047904	...	0.00354491	Fake	Fake
76	0.017	...	0.017	0.003388571	...	-0.000857143	True	True
100	0.021	...	0.028	0.005064516	...	0.00333871	True	True
110	-0.001	...	-0.08	-6.0606E-05	...	0.000848485	True	True
113	0.001	...	0.039	0	...	0.00125	True	True
114	0	...	-0.005	0.008646192	...	0.006066339	True	True
126	0.005	...	0.003	-0.00317778	...	0.024266667	True	True
130	0.019	...	0.267	-0.00046591	...	-0.001284091	True	True
131	0.024	...	0.022	0.01139777	...	0.008126394	True	True

어느 부분에서 정확도가 개선되었는지 좀 더
 엄밀히 살펴볼 필요가 있다고 판단하여 추가로
 가짜 뉴스 탐지 모델별 False Positive/Negative

분석과 가짜 뉴스 탐지 모델 간 정답 예측 빈도
 분석을 수행하였다.

<Figure 9>는 각 가짜 뉴스 탐지 모델별



<Figure 8> Accuracy Comparison of Fake News Detection Model

False Positive/Negative 분석 결과를 나타내고 있다. <Figure 9>의 (c) 뉴스와 트위터를 활용한 가짜 뉴스 탐지 모델의 분석 결과, (a) 뉴스를 활용한 가짜 뉴스 탐지 모델에 비해 True를 Fake로 잘못 예측한 8건 중 7건, 즉 약 88%를 제대로 예측하였고, Fake를 True로 잘못 예측한 19건 중 10건, 즉 약 53%를 제대로 예측하였음을 확인할 수 있었다. 반면 (b) 트위터를 활용한 가짜 뉴스 탐지 모델에 비해 True를 Fake로 잘못 예측한 13건 중 11건, 즉 약 85%를 제대로 예측하였고, Fake를 True로 잘못 예측한 12건 중 3건, 즉 약 25%를 제대로 예측하였음을 확인

하였다.

<Figure 10>은 가짜 뉴스 탐지 모델 간 정답 예측 빈도 분석 결과를 나타내고 있다. <Figure 10>의 (a)는 뉴스 데이터만을 활용한 가짜 뉴스 탐지 모델과 트위터 데이터만을 활용한 가짜 뉴스 탐지 모델 간 정답 예측 빈도 분석 결과를 나타내고 있으며, (b)는 (a) 결과에서 뉴스와 트위터 데이터를 활용한 가짜 뉴스 탐지 모델이 정답을 제대로 예측한 빈도를 나타내고 있다. 그 결과, 뉴스 활용 가짜 뉴스 탐지 모델과 트위터 활용 가짜 뉴스 탐지 모델에서 모두 잘못 예측한 뉴스 기사 13건 중 8건의 뉴스, 즉 약

News	Fake	True
Fake	5	19
True	8	24

(a) News Model

Twitter	Fake	True
Fake	12	12
True	13	20

(b) Twitter Model

News+Twitter	Fake	True
Fake	15	9
True	2	31

(c) News and Twitter Combined Model

<Figure 9> False Positive/Negative Analysis Results By Fake News Detection Model

		Twitter	
		O	X
News	O	23	12
	X	14	13

(a) Correct Prediction Frequency between News Model and Twitter Model

		Twitter	O	X
		News+Twitter	O	X
News	O		22	11
	X		11	8

(b) News and Twitter Combined Model Frequency of Predictions Contrast with (a)

<Figure 10> Correct Prediction Frequency Analysis Results

62%의 뉴스 기사를 제대로 예측함을 확인하였다. 이처럼 뉴스와 트위터 각각을 활용한 모델과 비교했을 때, False Positive/Negative와 정답 예측 빈도의 개선 비율이 본 제안 방법론의 성능을 향상시킨 요인으로 판단된다.

5. 결 론

본 연구는 기존의 과평가 되어 있던 가짜 뉴스 탐지 방법론의 정확도에 대한 공정성을 확보하고, 뉴스 자체만으로 가짜 뉴스를 탐지하기 어렵다는 한계를 극복하기 위해 뉴스와 트위터 데이터를 활용한 가짜 뉴스 탐지 기법을 제안하였다. 제안 방법론은 뉴스 데이터를 구조화하여 클러스터링을 통해 이슈 그룹을 생성하고, 이슈 그룹 내 진위여부가 다른 뉴스 기사를 학습 및 검증 데이터로 분리함으로써 가짜 뉴스 탐지 모델 정확도의 공정성을 확보하였다. 또한 뉴스와 트위터 데이터를 활용한 제안 방법론의 정확도가 뉴스와 트위터를 각각 활용한 가짜 뉴스 탐지 모델에 비해 약 28%p 향상됨을 증명하였다.

제안 방법론은 다음의 측면에서 학술적, 실무적 차원의 기여를 갖는다. 우선 학술적 측면에서 뉴스 데이터 사이에 존재하는 이슈간 연결 관계를 차단함으로써 기존의 과평가 되어있던 정확도의 공정성을 확보할 수 있는 방안을 제안했다는 점에서 의의를 갖는다. 이는 다른 도메인의 예측 모델과 달리 상대적으로 매우 적은 수의 새로운 가짜 뉴스에 대한 진위여부를 판별해야하는 특징을 갖는 가짜 뉴스 탐지가 엄격한 환경에서 제대로 된 예측 모델을 만들 수 있는 계기를 제공한 새로운 시도로 인정

받을 수 있다고 판단된다. 또한 기존의 뉴스 자체만으로 가짜 뉴스를 탐지하는 한계를 극복하고 트위터 데이터를 활용함으로써 보다 정확하게 가짜 뉴스 탐지를 가능하게 했다는 점에서 그 기여를 인정받을 수 있다. 본 연구의 성과는 실무적 측면에서 더욱 크게 나타날 것으로 기대한다. 가짜 뉴스 생성 비용과 이를 통해 거둘 수 있는 이익을 비교할 때, 가짜 뉴스 생성을 통해 부당한 방식으로 이익을 취하려는 시도는 더욱 증가할 것으로 예상된다. 또한 이에 따른 개인, 기업, 사회의 피해액도 이와 비례하여 증가할 것이므로, 가짜 뉴스를 사전에 탐지하여 이러한 피해에 선제적으로 대응하기 위한 시도가 반드시 이루어져야 한다. 특히 뉴스를 제공하는 언론사의 경우 가짜 뉴스의 양산은 해당 언론의 신뢰에 치명적인 영향을 미치므로, 제공한 뉴스의 팩트 체크에 대한 필요성을 크게 가질 수밖에 없다. 이러한 언론 기관들은 자사의 뉴스 기사뿐 아니라 이와 관련된 트윗으로부터 도출된 백터를 활용하는 제안 방법론을 적용함으로써, 이미 자체적으로 운영하고 있는 비기술적 접근 기반의 팩트 체크 서비스의 정확성과 효율성을 더욱 높일 수 있을 것이다.

하지만 본 연구는 향후 다음의 측면에서 보완이 필요하다. 본 연구는 현실적으로 데이터 수집의 어려움이 크기 때문에 매우 적은 수의 데이터를 사용하여 실험을 진행하였다. 제안 방법론의 성능이 향상됨을 보인 것은 매우 고무적인 현상이나, 그 폭이 너무 크기 때문에 어느 부분에서 정확도 향상에 기여를 했는지 좀 더 엄밀히 살펴볼 필요가 있다. 또한 제안 방법론 전 과정에 대한 정교화가 필요할 것으로 보이며, 향후 다른 가짜 뉴스 탐지 기법들과의 정확도 비교를 통한 추가 검증 과정이 필요할 것으로 판단된다. 더

나아가 트위터 뿐만 아니라 보다 다양한 소셜 데이터를 활용하여 가짜 뉴스 탐지 모델을 확장 역시 필요가 있을 것으로 보인다.

References

- [1] Albright, R., Taming Text with the SVD, SAS Institute Inc., 2006.
- [2] Chen, C., Wu K., Srinivasan V., and Zhang, X., “Battling the Internet Water Army: Detection of Hidden Paid Posters,” In Advances in Social Networks Analysis and Mining (ASONAM), 2013 IEEE/ACM International Conference, pp. 116-120, 2013.
- [3] Chen, Y., Conroy, N. J., and Rubin, V. L., “Misleading Online Content: Recognizing Clickbait as False News,” In Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection, pp. 15-19, 2015.
- [4] Conroy, N. J., Rubin, V. L., and Chen, Y., “Automatic Deception Detection: Methods for Finding Fake News,” Proceedings of the Association for Information Science and Technology, Vol. 52, No. 1, pp. 1-4, 2016.
- [5] Granik, M. and Mesyura, V., “Fake News Detection Using Naive Bayes Classifier,” In Electrical and Computer Engineering (UKRCON), 2017 IEEE First Ukraine Conference, pp. 900-903, 2017.
- [6] Hwang, Y. and Kwon, O., “A Study on the Conceptualization and Regulation Measures on Fake News: Focused on Self-Regulation of Internet Service Providers,” Journal of Media Law, Ethics and Policy Research, Vol. 16, No. 1, pp. 53-101, 2017.
- [7] Jin, Z., Cao, J., Jiang, Y. G., and Zhang, Y., “News Credibility Evaluation on Microblog with a Hierarchical Propagation Model,” In Data Mining (ICDM), 2014 IEEE International Conference, pp. 230-239, 2014.
- [8] Kim, D. J., “Semantic Analysis on Fake News through Portal Site and Social Network,” Master Thesis, 2017.
- [9] Kim, H. Y., “An Exploratory Study on Fake News Using Topic Modeling: Focused on Fake News Published in the Online Journalism,” Master Thesis, 2017.
- [10] Kwon, M., Jun, Y. W., and Im, H., “Controversy and Guideline Suggestion Surrounding Fake News in the Digital Media Age,” Journal of Korea Multimedia Society, Vol. 18, No. 11, pp. 1419-1426, 2015.
- [11] Kwon, S., Cha, M., Jung, K., Chen, W., and Wang, Y., “Prominent Features of Rumor Propagation in Online Social Media,” In Data Mining (ICEM), 2013 IEEE 13th International Conference, pp. 1103-1108, 2013.
- [12] Oh, S. U., “Current States and Limitations of Automated Fact Checking Technology,” Journal of Cybercommunication Academic Society, Vol. 34, No. 3, pp. 137-180, 2017.

- [13] Park, J. H. and Kim, Y. I., "Development of a Fake News Discrimination System using SVM Classifier," Proceedings of KIIT Summer Conference, pp. 354-355, 2017.
- [14] Rubin, V. L., Chen, Y., and Conroy, N. J., "Deception Detection for News: Three Types of Fakes," Proceedings of the Association for Information Science and Technology, Vol. 52, No. 1, pp. 1-4, 2016.
- [15] Lee, D., Kim, Y., and Kim, K., "Topic Based Hierarchical Network Analysis for Entrepreneur Using Text Mining," The Journal of Society for e-Business Studies, Vol. 23, No. 3, pp. 33-49, 2018.
- [16] Lee, S. and Kim, H. J., "Keyword Extraction from News Corpus using Modified TF-IDF," The Journal of Society for e-Business Studies, Vol. 14, No. 4, pp. 59-73, 2009.
- [17] Salton, G., Wong, A., and Yang, C. S., "A Vector Space Model for Automatic Indexing," Communications of the ACM, Vol. 18, No. 11, pp. 613-620, 1975.
- [18] Sethi, R. J., "Spotting Fake News: A Social Argumentation Framework for Scrutinizing Alternative Facts," In Web Services (ICWS), 2017 IEEE International Conference, pp. 866-869, 2017.
- [19] Weiss, S. M., Indurkha, N., and Zhang, T., Fundamentals of Predictive Text Mining, Springer, 2010.

저 자 소개



현윤진

2013년

2015년

2015년~현재

(E-mail: yoonjin0630@kookmin.ac.kr)

국민대학교 경영정보학 (학사)

국민대학교 비즈니스IT전문대학원 경영정보학 (석사)

국민대학교 비즈니스IT전문대학원 경영정보학 박사과정



김남규

1998년

2000년

2007년

2007년~현재

(E-mail: ngkim@kookmin.ac.kr)

서울대학교 컴퓨터공학 (학사)

한국과학기술원 경영공학 (석사)

한국과학기술원 경영공학 (박사)

국민대학교 경영정보학부 교수