「정보시스템연구」 제32권 제2호
한국정보시스템학회
2023년 6월, pp. 87~108

http://dx.doi.org/10.5859/KAIS.2023.32.2.87

영상정보를 활용한 소셜 미디어상에서의 가짜 뉴스 탐지: 유튜브를 중심으로*

장윤호** · 최병구***

〈목 차〉

Ⅰ. 서 론

지난 몇 년간 유튜브, 카카오톡, 페이스북 등과 같은 소셜 미디어를 통한 뉴스 소비가 빠르게 증가하고 있다. 미국의 퓨리서치센터(Pew Research Center)의 2020년 조사에 따르면 미국인의 53%가 소셜 미디어를 뉴스 소비 경로로 활용하고 있는 것으로 나타났다(Shearer, 2021). 한국언론진흥재단의 조사 역시 미국 및 한국을 포함한 46개국 국민의 39%가 뉴스 소비 경로로 소셜 미디어를 이용하고 있는 것으로 나타났다(최진호, 박영흠, 2022). 이처럼 뉴스 소비의 경로로 소셜 미디어가 각광 받고 있는 이유는 사용자가 소셜 미디어를 통해 본인의 기호에 부합하는 뉴스를 더욱 쉽게 찾을 수 있을 뿐 아니라 이를 다른 사람과 빠르게 공유할 수 있기 때문이다(한혜주, 이경미, 2014).

그러나 소셜 미디어를 활용한 뉴스 소비는 TV나 신문을 통한 전통적 뉴스 소비에 비해 상대적으로 가짜 뉴스에 더 취약한 것도 사실이다(Luvembe et al., 2023). 소셜 미디어의 경우 뉴스 유통 비용 감소와 정보 과잉 문제 해결을 위해 사용자에게 최적의 정보를 제공하는 서비스를 지원하고 있다. 그러나 이러한 서비스는

* 이 논문은 2021년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2021S1A5A2A01068368).
** 바이브컴퍼니 S.C.I연구소 연구원, yhchang1120@naver.com(주저자)
*** 국민대학교 AI빅데이터융합경영학과 교수, h2choi@kookmin.ac.kr(교신저자)

사용자의 뉴스 접근성 및 선택권을 제한하고 뉴스 소비에 있어 소셜미디어 시스템에 의존하는 '필터 버블(filter bubble)[1]'(Pariser, 2011) 문제를 야기한다. 따라서 소셜 미디어를 통해 가짜 뉴스를 접하게 되면 이와 유사한 가짜 뉴스에 더욱 노출되어 고착화되는 확증 편향의 위험성이 발생할 가능성이 매우 큰 것이 사실이다. 또한, 소셜 미디어는 그 특성상 다수의 사용자에게 실시간으로 유통되기 때문에 출처가 불분명한 뉴스가 검증 절차를 거치지 않고 빠르게 유통될 가능성이 매우 크다. Vosoughi et al.(2018)의 연구에 따르면 가짜 뉴스의 온라인 확산속도가 진짜 뉴스에 비해 6배 빠르며 그 유통 범위 또한 진짜 뉴스보다 35% 넓은 것으로 조사되었다. 더욱이 가짜 뉴스는 진짜 뉴스보다 더 많은 관심을 받는 것으로 나타났다. Silverman(2016)에 의하면 2016년 미국 대선 전 3개월간 페이스북상에서 가장 많이 공유되었던 20개 진짜 뉴스는 730만 번의 반응(공유, 댓글 등)을 보인 반면 상위 20개 가짜 뉴스는 870만 번의 반응을 보인 것으로 조사되었다.

가짜 뉴스는 정치·경제적 목적을 위해 생산되며 목적 달성 과정에서 개인, 기업, 사회 등에 경제적 손실을 포함한 막대한 사회적 비용을 발생시킨다는 점에서 매우 심각한 문제이다(Capuano et al., 2023). 정민과 백다미(2017)는 전체 뉴스 가운데 가짜 뉴스 건수가 1% 정도 유포되는 경우 개인, 기업, 사회에 발생하는 경제적 비용이 연간 약 30조 900억 원에 이른다고 주장하였다. 가짜 뉴스는 또한 집단의 양극화와 극단주의를 부추겨 사회적 갈등을 증가시키고 통합을 저해한다. 예를 들면, 그리스에서는 여행 제한이 풀렸다는 가짜 뉴스로 인해 그리스 경찰과 이민자 간의 대규모 유혈사태가 발생하였다(Guo et al., 2020).

이러한 문제점을 해결하기 위해 가짜 뉴스 탐지와 관련된 다양한 연구가 진행되어 왔으며 가짜 뉴스 탐지와 관련된 우리의 이해를 일정 정도 증진시킨 것도 사실이다. 그러나 가짜 뉴스 탐지와 관련된 기존 연구는 텍스트 데이터만을 주로 사용하고 있다는 점, 영상정보 위주의 최신 뉴스 소비 추세를 반영하고 있지 못하다는 점, 그리고 영상정보를 활용한 매우 소수의 연구조차 대단히 제한적인 영상정보만을 사용하고 있지 못하다는 점 등이 단점으로 지적되고 있다.

텍스트 및 이미지 중심 기존 연구들이 영상정보 위주의 최신 뉴스 소비 추세를 반영하지 못하고 있는 한계점을 극복하기 위하여 본 연구에서는 텍스트 정보뿐만 아니라 영상정보를 동시에 고려한 가짜 뉴스 탐지 방법론을 제안함으로써 가짜 뉴스 탐지 성능을 개선하고자 한다. 특히 뉴스 영상뿐만 아니라 이와 관련한 다양한 영상정보를 활용함으로써 진정한 의미의 영상정보 기반 가짜 뉴스 탐지를 수행하고자 한다. 나아가 가짜 뉴스와 관련한 다양한 특성 조합(feature combination)의 학습을 통해 가짜 뉴스 탐지에 가장 적합한 특성 조합을 확인함으로써 기존 연구의 한계를 극복하고자 한다.

이러한 목적을 달성하기 위해 본 연구에서는

1) 필터 버블(filter bubble)은 개인화된 검색 알고리즘에 의해 선택적 검색 결과에 노출된 사용자가 자신의 관점에 동의하지 않는 정보로부터 분리되어 제한적이고 맞춤화된 세계관을 갖게 되는 지적 고립 상태를 의미한다(Pariser, 2011).

진정한 의미의 뉴스 영상인 뉴스를 진행하는 화자의 얼굴 표정을 점수화하고 이를 데이터로 입력하여 가짜 뉴스 탐지를 시도하고자 한다. 이와 함께 영상 총 조회수, 영상 길이, 영상 설명 단어 등 영상과 관련한 다양한 메타 데이터를 활용함으로써 가짜 뉴스 탐지의 성능을 개선하고자 한다. 본 연구의 주요 분석 대상인 영상정보 기반의 뉴스는 텍스트 정보인 뉴스 대본, 뉴스를 전달하는 화자의 표정, 영상과 관련한 메타 데이터 등과 같은 다양한 특성을 포함하고 있다. 따라서 단순히 영상정보를 활용하는 것이 중요한 것이 아니라 영상정보로부터 추출된 관련 특성을 어떻게 조합하느냐에 따라 가짜 뉴스 탐지 성능이 변화할 가능성이 매우 크다(Cao et al., 2018). 따라서 광범위한 특성을 추출하고 효과적인 가짜 뉴스 탐지를 위한 특성 간 조합을 확인하고자 한다. 본 연구는 '뉴스 읽기'에서 영상정보 기반의 '뉴스 보기'로 점차 변화하고 있는 최근 뉴스 소비 형태를 반영함으로써 가짜 뉴스와 관련된 텍스트 위주의 기존 연구를 보완할 수 있을 것이다.

본 연구는 다음과 같이 구성된다. 다음 장에서는 가짜 뉴스 탐지를 위한 기존 연구들을 요약한다. 제3장에서는 연구 모형, 데이터 셋(data set), 평가방법 등을 포함한 연구 방법을 설명한다. 제4장에서는 데이터를 분석하고 분석결과와 이의 함의를 논의한다. 마지막 장에서는 본 연구의 결론 및 향후 연구과제를 제안한다.

Ⅱ. 선행 연구

지금까지 가짜 뉴스와 관련하여 가짜 뉴스의 정의와 개념(Gelfert, 2018), 법 규제(Jang & Kim, 2018), 가짜 뉴스의 유통 및 소비 패턴(Vosoughi et al., 2018) 등 다양한 학문 영역에서 가짜 뉴스와 관련된 연구가 이루어져 왔다. 특히 가짜 뉴스로 인한 피해를 사전에 막기 위한 방편으로써 가짜 뉴스 탐지가 주목받고 있으며 이와 관련한 다양한 연구가 진행되어 왔다. 기존 가짜 뉴스 탐지 연구는 크게 컨텍스트 기반(context-based)과 콘텐츠 기반(contents-based) 연구로 구분할 수 있다(Bondielli et al., 2019; Shu et al., 2017). 컨텍스트 기반 연구는 사용자의 게시글 수, 팔로워 수 등과 같은 소셜 미디어 사용자의 참여 행위와 관련된 특성 또는 사용자 간 연결 관계를 기반으로 한 네트워크 특성에 초점을 둔 연구 방법이다(Kwon et al., 2013). 이러한 연구들은 가짜 뉴스의 확산 패턴을 파악할 수 있는 장점이 있으나 개인정보 보호로 인한 제약 때문에 데이터의 수집이 어려워 상대적으로 많은 연구가 이루어지지 않았다(Bondielli et al., 2019). 이에 비해 콘텐츠 기반 연구는 뉴스 텍스트 분석을 통한 특성 추출(feature extraction)과 머신러닝이나 딥러닝 기반 분류 모델을 결합하는 연구 방법이다(민진영, 이애리, 2021; 신동훈 등, 2022). 이 방법은 가짜 뉴스의 확산 패턴을 파악하기 어렵다는 한계점이 있지만, 상대적으로 데이터 수집이 용이하고 뉴스 콘텐츠 중심의 분석이 가능하다는 점에서 많은 연구자들이 활용하고 있다(Jarraji & Safari, 2023). 따라서 본 연구 역시 뉴스 콘텐츠를 기반으로 한 연구를 진행하였다.

콘텐츠 기반 연구는 가짜 뉴스 탐지를 위한 분석에 사용한 특성의 유형을 기준으로 i) 텍스

트 활용, ii) 이미지 활용, iii) 영상 메타 데이터 중심 연구의 세 가지 범주로 구분할 수 있다. 첫 번째 범주에 속한 연구들은 가짜 뉴스 탐지에 있어 텍스트 길이, 텍스트 감성, 텍스트의 문체적 특징 등과 같은 텍스트 정보 활용에 초점을 두고 있다(Hu et al., 2014; Ito et al., 2015). 예를 들면, Ajao et al.(2019)은 소셜 네트워크 서비스에서 작성된 메시지가 가짜 뉴스 확산의 원인임을 지적하고 가짜 뉴스와 가짜 뉴스가 아닌 트윗 간의 감성 점수 차이가 존재할 것이라는 가설을 제시하였다. 가설 검증을 위해 긍정적 단어와 부정적 단어의 비율을 새로운 특성으로 제시하고 해당 비율을 감성 점수로 활용하여 분석을 시도하였다. 분석결과 가짜 뉴스와 가짜 뉴스가 아닌 트윗에 따라 감성 점수의 차이가 존재하는 것을 확인하고 감성을 고려함으로써 가짜 뉴스 탐지 성능이 개선됨을 실증하였다. 나아가 감성을 고려하지 않은 가짜 뉴스 탐지 연구와 비교하여 감성을 고려하였을 때 가짜 뉴스 탐지 성능이 개선됨을 실증하였다. Bhutani et al.(2019)은 가짜 뉴스 분류가 특정 주제에 대한 작성자의 태도에 의존하기 때문에 감성을 고려한 연구가 필요함을 주장하고 단어 빈도-역문서 빈도 (TF-IDF: term frequency-inverse document frequency)와 감성 점수를 활용한 가짜 뉴스 탐지 방법론을 제안하고 이를 세 가지 뉴스 데이터셋(Politifact, Kaggle, Emergent data set)에 적용하여 감성을 고려함으로써 가짜 뉴스 탐지의 정확도가 향상됨을 검증하였다. 정호선(2019)은 단어 기반의 특징을 사용한 기존의 단어 빈도-역문서 빈도와 워드 임베딩 방법이 가짜와 진짜 텍스트가 가지는 문체적 특징을 반영하지 못한다는 한계점을 지적하며, 문체적 특징을 반영하는 어휘 변수, 구문론적 변수, 심리언어학적 변수를 사용하여 전통적인 단어 빈도-역문서 빈도 기반의 모델보다 우수한 성능을 검증하였다. 현윤진과 김남규(2018)는 가짜 뉴스의 조작 정밀도가 높을수록 뉴스 자체에 대한 분석만으로는 진위 여부를 식별하기 어렵다는 한계를 지적하였다. 이를 극복하기 위해 뉴스와 트위터 데이터를 함께 사용하여 가짜 뉴스 탐지를 시도하였다. 연구결과 뉴스와 트위터를 함께 사용한 경우의 가짜 뉴스 탐지 정확도가 뉴스와 트위터를 개별적으로 사용했을 때보다 우수함을 실증하였다. 이외에도 Potthast et al.(2017)은 다양한 작문 스타일(writing style)을 기준으로, Horne & Adali(2017)는 제목 스타일(title style)을 기준으로 가짜 뉴스 탐지를 시도하였다.

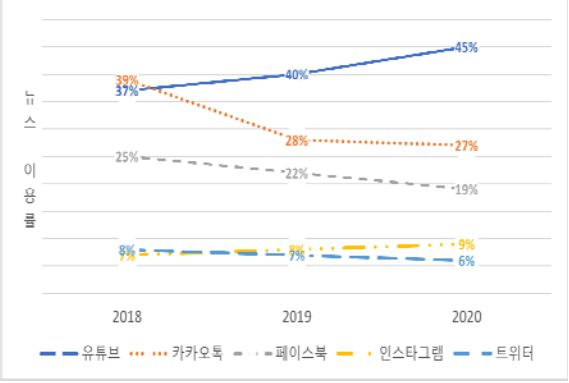
두 번째 범주에 속한 연구들은 가짜 뉴스 탐지에 있어 이미지 정보의 중요성을 지적하고 이미지 진위 여부, 이미지 색상 비율, 사이즈, 이미지 토픽 등과 같은 다양한 이미지의 속성을 활용하여 가짜 뉴스 탐지를 시도하였다. 예를 들면, Masciari et al.(2020)은 위조된 이미지가 가짜 뉴스의 주요 원천으로 사용되는 것에 주목하여 이미지 진위 여부 확인을 통한 가짜 뉴스 탐지를 시도하였다. 그들은 딥러닝 기반의 가짜 이미지 분류기를 개발하고 이를 활용하여 뉴스에 포함된 이미지 진위 여부를 새로운 특성으로 제시하였다. Singh et al.(2021)은 텍스트 정보와 시각적 정보를 모두 포함하는 뉴스 기사가 증가함에 따라 텍스트와 이미지 정보의 통합 분석 필요성을 제시하였다. 이를 위해 뉴스 기사와 이미지가 가지는 내용, 구성, 감성, 조작의 속성을 특성으로 사용하고 이를 통해

가짜 뉴스 탐지 모형의 성능 개선이 가능함을 검증하였다. 이외에도 Fridrich & Kodovsky (2012)는 인접한 이미지 픽셀의 노이즈 패턴(noise patterns)을, Huh et al.(2018)은 교환 가능한 이미지 파일 포맷(exchangeable image file format)의 메타 데이터를 활용하여 가짜 뉴스 탐지를 시도하였다.

마지막 범주에 속한 연구들은 영상 메타 데이터를 활용하여 가짜 뉴스 탐지를 시도한 연구로 매우 소수의 연구만이 이루어져 왔다. 예를 들면, Serrano et al.(2020)은 가짜 뉴스의 주요 원천으로 유튜브의 심각성을 지적하며 영상정보를 활용한 가짜 뉴스 탐지 연구를 제안하였다. 이를 위해 영상 코멘트 가운데 음모론적 코멘트의 비율을 새로운 특성으로 제시하고 이를 활용함으로써 가짜 뉴스 탐지의 성능이 개선 가능함을 주장하였다.

기존 연구를 통합 분석해 보면 다음과 같은 흥미로운 사실을 발견할 수 있다. 첫째, 대다수의 연구는 가짜 뉴스 탐지에 있어 텍스트 데이터만을 활용하고 있다(Ajao et al., 2019; Bhutani et al., 2019). 최근 소셜 미디어의 사용이 급증함에 따라 텍스트와 이미지를 결합한 뉴스 기사가 증가하고 있다. 이는 이미지 정보가 텍스트 정보에 비해 해당 뉴스를 더욱 생생하게 묘사하여 사용자들에게 보다 깊은 인상을 줄 수 있기 때문이다(Guo et al. 2020). 나아가 가짜 뉴스의 주요 원천으로 위조된 이미지의 사용이 점점 증가하고 있는 추세이다(Masciari et al., 2020). 이러한 현실 변화에 비추어 볼 때 텍스트에만 의존하는 기존 연구들은 가짜 뉴스 탐지에 있어 많은 한계점이 있다.

둘째, 텍스트와 이미지를 결합하여 가짜 뉴스 탐지를 시도한 연구의 경우 영상 위주의 최신 뉴스 소비 추세를 반영하지 못하고 있다. 많은 연구들은 가짜 뉴스 탐지에 있어 텍스트 위주의 한계점을 극복하고자 이미지 정보를 함께 고려하였다. 예를 들면, Jin et al.(2016)은 가짜 뉴스와 진짜 뉴스 간 이미지 정보의 통계적 특성이 다름을 주장하고 이를 판별하기 위한 지표로 시각적 일관성 점수(visual coherence score)와 다양성 점수(visual diversity score)를 제시하여 가짜 뉴스 탐지를 시도하였다. 그러나 텍스트와 이미지 정보를 결합하여 가짜 뉴스



<그림 1> 소셜 미디어별 뉴스 소비 방식의 변화 추세
(출처: 박아란, 이소은, 2020)

탐지를 시도한 연구들은 영상 위주의 최신 뉴스 소비 추세를 반영하지 못하고 있다. 최근 조사에 따르면 뉴스 소비에 있어 페이스북, 트위터, 카카오톡 등과 같은 텍스트 기반 소셜 미디어의 사용은 점점 감소하고 있는 반면 유튜브와 같은 영상 기반 소셜 미디어의 활용은 점점 증가하고 있다(박아란, 이소은, 2020). 예를 들면, 우리나라의 경우 카카오톡을 통한 뉴스 소비는 2018년 39%, 2019년 28%, 2020년 27%로 점차 감소하고 있는 반면 유튜브를 통한 뉴스 소비는 2018년 37%, 2019년 40%, 2020년 45%로 점점 증가하고 있다(<그림 1> 참조). 이처럼 뉴스 소비 방식이 텍스트 기반에서 영상 기반으로 전환되고 있음에도 불구하고 가짜 뉴스 탐지에 있어 영상정보를 활용한 연구는 거의 이루어지지 않고 있다.

셋째, 많은 연구들이 가짜 뉴스 탐지에 있어 영상정보의 중요성을 강조하고 있으나(Guo et al. 2020), 매우 소수의 연구만이 가짜 뉴스 탐지를 위해 영상정보를 활용하고 있다. 그러나 이러한 연구조차 다양한 영상정보 가운데 극히 일부분만을 사용하고 있어 영상정보를 충분히 활용하지 못하고 있다. 예를 들면, 영상정보를 활용한 Serrano et al.(2020)의 연구조차 영상 코멘트(comment)만을 사용하였다는 점에서 진정한 의미의 영상정보를 활용한 가짜 뉴스 탐지 연구라 보기 어렵다. 이로 인해 아직 최신 영상정보 기반 뉴스 소비 추세를 반영한 정교한 가짜 뉴스 탐지가 이루어지지 못하고 있는 실정이다.

Ⅲ. 연구 방법

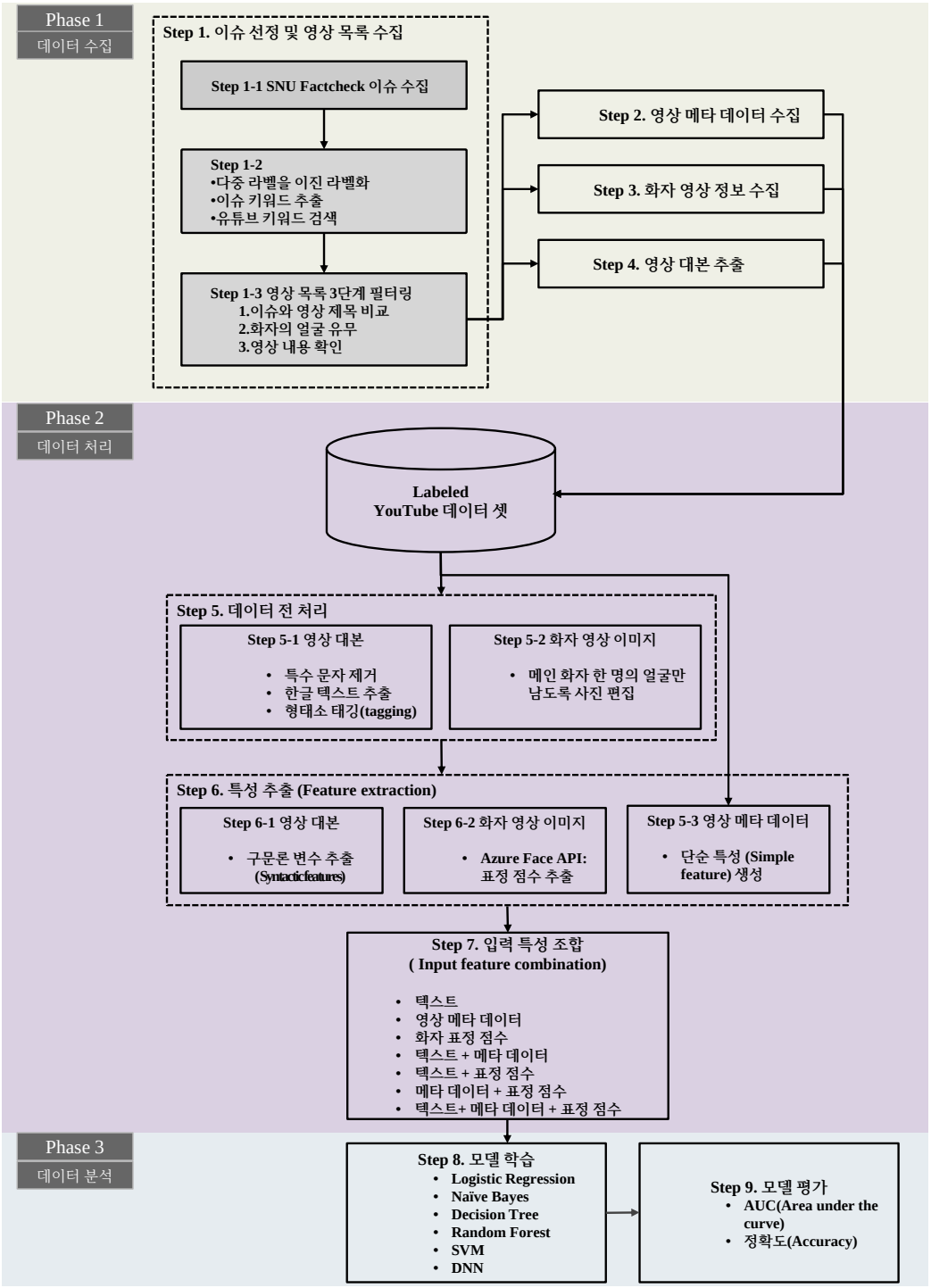
영상정보 활용을 통한 가짜 뉴스 탐지 성능 개선이라는 본 연구의 목적을 달성하기 위한 절차를 요약하면 다음 <그림 2>와 같다. 본 연구의 절차는 크게 학습과 검증에 필요한 데이터 수집, 수집된 데이터로부터 특성을 추출하는 처리, 머신러닝과 딥러닝 알고리즘을 통해 가짜 뉴스를 판별하는 분석의 3단계로 구성된다.

3.1 데이터 수집

가짜 뉴스 탐지 모형의 학습과 검증에 사용할 데이터는 이슈 및 관련 유튜브 영상 목록 수집, 영상 목록을 기반으로 영상 메타 데이터 수집, 영상 속 화자의 얼굴 이미지 수집, 영상 대본 및 자막(즉, 뉴스 텍스트 정보) 추출의 4단계를 거쳐 수집되었다.

이슈 및 관련 유튜브 영상 목록 수집 단계에서는 먼저 서울대학교 언론정보연구소에서 운영하는 SNU 팩트체크 서비스[2]를 이용하여 2017년 3월부터 2021년 8월까지 등록된 정치 이슈를 1, 123개를 수집하였다(<그림 2>의 step 1-1 참조). 수집된 이슈의 예로는 “대통령이 미국 방문에서 푸대접 받았다.”, “종전 선언하면 군대 안가도 된다.” 등이 있다. 각 이슈는 언론사의 검증 결과에 따라 ‘전혀 사실 아님’, ‘대체로 사실 아님’, ‘절반의 사실’, ‘대체로 사실’, ‘사실’, ‘판단 유보’로 라벨링 되어 제공되지만

2) 다른 팩트체크 서비스와 달리 SNU 팩트체크는 i) 2017년부터 현재까지 꾸준히 운영되고 있으며, ii) 30개 언론사의 참여로 이슈에 대한 교차 검증이 이루어질 뿐 아니라, 3) 정치, 경제, 사회, 문화 등 제반 분야에서 사실 검증이 필요하다고 판단되는 공적 사안에 대한 검증 서비스를 제공하고 있어 본 연구의 목적에 매우 적합한 정보를 제공하고 있다.


Phase 1
데이터 수집
Step 1. 이슈 선정 및 영상 목록 수집
Step 1-1 SNU Factcheck 이슈 수집
Step 1-2
•다중 라벨을 이진 라벨화
•이슈 키워드 추출
•유튜브 키워드 검색
Step 1-3 영상 목록 3단계 필터링
1.이슈와 영상 제목 비교
2.화자의 얼굴 유무
3.영상 내용 확인
Step 2. 영상 메타 데이터 수집
Step 3. 화자 영상 정보 수집
Step 4. 영상 대본 추출
Phase 2
데이터 처리
Labeled
YouTube 데이터 셋
Step 5. 데이터 전 처리
Step 5-1 영상 대본
• 특수 문자 제거
• 한글 텍스트 추출
• 형태소 태깅(tagging)
Step 5-2 화자 영상 이미지
• 메인 화자 한 명의 얼굴만 남도록 사진 편집
Step 6. 특성 추출 (Feature extraction)
Step 6-1 영상 대본
• 구문론 변수 추출 (Syntacticfeatures)
Step 6-2 화자 영상 이미지
• Azure Face API: 표정 점수 추출
Step 5-3 영상 메타 데이터
• 단순 특성 (Simple feature) 생성
Step 7. 입력 특성 조합
(Input feature combination)
• 텍스트
• 영상 메타 데이터
• 화자 표정 점수
• 텍스트 + 메타 데이터
• 텍스트 + 표정 점수
• 메타 데이터 + 표정 점수
• 텍스트+ 메타 데이터 + 표정 점수
Phase 3
데이터 분석
Step 8. 모델 학습
• Logistic Regression
• Naïve Bayes
• Decision Tree
• Random Forest
• SVM
• DNN
Step 9. 모델 평가
• AUC(Area under the curve)
• 정확도(Accuracy)

<그림 2> 연구 절차

본 연구에서는 분석의 용이성을 위해 ‘대체로 사실’과 ‘사실’은 참(true)으로, ‘전혀 사실 아님’, ‘대체로 사실 아님’, ‘절반의 사실’은 거짓(false)으로 구분하였으며 73개의 ‘판단 유보’ 이슈는 분석에서 제외하였다. 그 결과 1,050개의 라벨링된 이슈를 확보하였다.

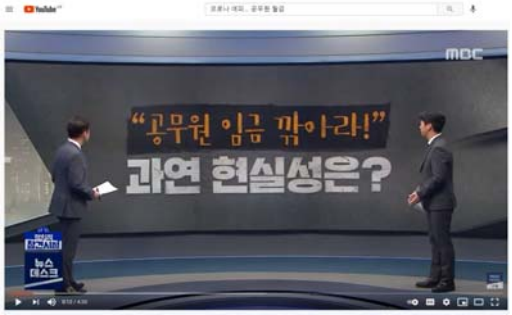
확보된 이슈 관련 영상 목록을 수집하기 위해 Python KoNLPy 패키지를 사용하여 수집된 이슈 텍스트로부터 [대통령, 미국, 방문, 푸대접], [종전, 선언, 군대] 등과 같은 키워드를 추출하였다. 이렇게 추출된 키워드를 사용하여 유튜브로부터 관련 영상을 검색하여 수집하였다(<그림 2>의 step 1-2 참조). 검색 결과 목록의 하위로 갈수록 키워드와의 연관성이 점점 떨어지기 때문에 상위 5개의 영상 URL만을 사용하였으며 그 결과 5,250개의 영상 목록을 수집하였다. 입력 데이터의 품질을 높이기 위해 수집된 영상 목록은 총 3단계 필터링(filtering) 작업을 통해 정제되었다(<그림 2>의 step 1-3 참조). 먼저 해당 영상의 제목과 이슈 텍스트만을 비교하여 관련성 여부를 확인하였다. 이슈 텍스트와 수집된 관련 영상 목록의 예시는 다음 <표 1>과 같다.

비교 그 결과 영상과 이슈가 관련이 없는 4,181개 영상을 제외하였다. 예를 들면, <그림 3>과 같이 <표 1>에 나타나 있는 “차량집회 막는 나라는 없다”라는 이슈의 관련 동영상으로 검색된 “광복절 집회 열었던 보수단체, 기자회견 열어... 이 시각 개천절 집회 상황/YTN”의 경우 차량집회라는 이슈와 관련 없는 보수단체

<표 1> 이슈 텍스트와 영상 제목 비교 예시

이슈 ID	이슈 텍스트	영상 제목	비교 결과
...	...	...	...
snufc#20	“백선영 장군의 묘가 파헤쳐졌다”	[팩트체크]“파헤쳐진 백선엽 장군의 묘?” 사진속의 진실은?/JTBC 뉴스룸	일치
snufc#20	“백선영 장군의 묘가 파헤쳐졌다”	한 달도 안돼 ‘백선엽 파묘’?	일치
...	...	...	...
snufc#20	“백선영 장군의 묘가 파헤쳐졌다”	백선엽 장군 안장식후 묘소에서 풍수해설: 수맥이 강해 우려된다	불일치
snufc#21	차량집회 막는 나라는 없다	[팩트의 무게]‘승차집회’제한하는 나라는 없다?(2020.09.25./뉴스테스크/MBC)	일치
snufc#21	차량집회 막는 나라는 없다	광복절 집회 열었던 보수단체, 기자회견 열어...이 시각 개천절 집회 현황/YTN	불일치
...	...	...	...
snufc#21	차량집회 막는 나라는 없다	“집회·기자회견 금지” 광화문 ‘원천봉쇄’...과잉 대응 논란/뉴스A	불일치
...	...	...	...
snufc#25	공무원 월급 지역화폐로 지급 가능하다?	코로나 여파...공무원 월급 지급도 바꿔!/안동MBC	일치
snufc#25	공무원 월급 지역화폐로 지급 가능하다?	[이슈&직설]코로나 직격탄 맞은 자영업자...살릴 방법은?	불일치
...	...	...	...
snufc#25	공무원 월급 지역화폐로 지급 가능하다?	“공무원 임금 깍아라!” 과연 현실성은? 코로나 재확산? “정부탓vs”집회탓“/뉴스테스크/MBC	일치
...	...	...	...

<그림 3> 이슈와 영상 간의 불일치



<그림 4> 화자 얼굴 영상 파악 불가능

의 광복절 집회 영상이 주를 이루었기 때문에 이후 분석에는 사용하지 않았다.

다음으로 이슈를 발언하는 화자의 얼굴이 영상에 나오는지를 파악하였다. 예를 들면, <그림 4>와 같이 이슈와 연관된 영상이기는 하나 화자의 얼굴이 해당 영상에 나오지 않는 데이터 638개를 제외하였다[3]. 마지막으로 2인의 연구자가 영상 내용을 직접 확인하여 이슈와 관련성이 없는 22개의 영상을 제외하였다. 이를 통해 최종적으로 참(true)으로 판명된 영상 303개와 거짓(false)으로 판명된 영상 409개를 분석 데이터로 선정하였다. 단계별 이슈 및 영상 개수를 요약하면 다음 <표 2>와 같다.

<표 2> 이슈 및 영상 선정 절차 및 개수

	단계	개수
이슈 선정	정치 이슈 수집(2017년 3월부터 2021년 8월까지 등록된 정치 이슈)	1,123
	• 판단 유보 이슈(참, 거짓이 불분명) 제외	73
	최종 선정 이슈	***1050***
영상 선정	정치 이슈별 영상(키워드 검색, 이슈별 5개 영상: 1050×5)	5,250
	• 확인된 이슈 관련 없는 영상 제외	4,181
	• 화자의 얼굴 식별 불가능 영상 제외	638
	– 전문가에 의해 확인된 이슈와 영상 간의 불일치 영상 제외	22
	최종 선정 영상	***409***

영상 메타 데이터는 영상의 기본 정보를 의미하며 채널명, 제목, 태그, 자막, 영상 설명 등 영상 고유의 정보를 포함한다. 이러한 메타 데이터는 주로 합성된 영상을 분류하는 가짜 영상 탐지 연구에 사용되어 왔다(Papadopoulou et al., 2017). 본 연구에서는 Python 셀레니움(selenium) 패키지를 사용하여 채널명, 영상 설명, 좋아요, 싫어요, 제목, 업로드 날짜, 영상 길이, 조회수, 코멘트, 코멘트 작성자 정보 등의 메타 데이터를 수집하였다 (<그림 2>의 step 2 참조).

본 연구는 텍스트 중심의 기존 연구와 달리 화자의 영상 이미지를 가짜 뉴스 탐지의 주요 특성으로 활용한다. 이를 위한 화자의 영상 이

3) <그림 4>는 위기 극복을 위해 공무원의 임금 삭감이라는 뉴스를 전달함으로써 <표 1>의 "공무원 월급 지역화폐로 지급 가능하다"라는 이슈와 연관성이 있다.

미지는 오픈소스 미디어 플레이어로 운영체제의 제한 없이 사용 가능하며 재생 중인 영상을 프레임 단위로 캡처하는 기능을 제공하고 있는 VLC 미디어 플레이어(media player)를 사용하여 수집하였다. 영상 이미지 수집을 위해 재생 속도는 20배로 지정하였으며 초당 30프레임 영상을 기준으로 약 4초에 1회씩 캡처를 수행하여 영상 이미지 정보를 수집하였다(<그림 2>의 step 3 참조). 앞서 단계에서 선정된 409개의 영상에 대해 총 14,274장의 이미지를 수집하였다. 이 가운데 이슈를 전달하는 화자의 얼굴이 정면을 바라보고 있지 않은 7,432장을 제외한 6,842장을 분석을 위한 최종 데이터로 선정하였다.

유튜브 뉴스 영상의 경우 영상 대본을 수집할 수 있는 영상과 그렇지 않은 영상으로 구분된다. 영상 대본이 있는 경우는 언론사 페이지에 게시된 영상 대본을 수집하여 활용하였으며 영상 대본이 없는 경우는 Python youtube-transcript-API 패키지를 사용하여 자동 생성된 자막을 추출하여 영상 대본으로 활용하였다(<그림 2>의 step 4 참조).

3.2 데이터 처리

데이터 처리는 전처리(preprocessing), (영상 대본, 화자의 영상 이미지, 영상 메타 데이터로부터) 특성 추출, 추출된 특성의 조합을 포함한 3단계로 구성된다. 데이터 분석을 위해서는 수집된 텍스트 및 영상 이미지 데이터 셋에 대한 전처리가 필요하다. 영상 대본인 텍스트 데이터에 대한 전처리를 위해 먼저 특수 문자와 문장 부호를 제거한 후 문자열의 모든 공백을 제거하였다. 다음으로 PyKoSpacing 패키지를 사용하여 자동 띄어쓰기를 수행한 후 py-hanspell 패키지를 사용하여 자동 맞춤법 검사를 수행하였다. 마지막으로 구문론 변수 추출을 위해 KoNLPy 패키지의 Komoran 클래스를 사용하여 형태소를 태깅(tagging)하였다(<그림 2>의 step 5-1 참조).

가짜 뉴스 탐지를 위한 영상 이미지의 경우 이슈에 대해 발언하는 화자 한 명의 얼굴 이미지만 존재해야 한다. 기자와 앵커 또는 앵커와 뉴스 관련인 등 불특정 다수의 인물이 한 화면에 동시에 나타나는 경우 분석 대상이 되는 화자를 식별할 수 없기 때문이다. 따라서 수작업을 통해 화자 한 명의 얼굴 이미지만 존재하도록 영상 이미지 데이터에 대한 전처리를 위한 이미지 편집을 수행하였다(<그림 2>의 step 5-2 참조).

수집된 텍스트 데이터로부터 특성을 추출하기 위하여 품사의 출현 빈도를 기반으로 구문론 특성 추출 방식을 활용하였다(<그림 2>의 step 6-1 참조). 수집된 텍스트 데이터 셋의 많은 부분은 자동 생성된 자막으로 이루어져 있다. 자동 생성된 자막은 음성인식 기술을 통해 생성되기 때문에 영상 속 소음, 화자의 발음, 말투에 따라 잘못 인식될 가능성이 매우 크며 이러한 문제점으로 인해 토픽 모델링, 단어 빈도-역 문서 빈도, 워드2벡터(word2vec)를 사용하여 그 특성을 추출하는 데 한계가 있다. 형태소 태깅을 활용한 구문론 특성 추출 방식을 활용함으로써 이러한 한계점을 극복할 수 있다(정호선, 2019). 기존 연구를 기반으로(Perez-Rosas et al., 2017; Zubiaga et al., 2016) 본 연구에서는 텍스트 데이터 셋의 형태소 태그 중

<표 3> 텍스트 데이터 특성

특성 이름	설명	특성 이름	설명
noun_count	일반명사와 고유명사 개수	adverb_count	일반부사와 접속부사 개수
adj_count	형용사 개수	neg_copula_count	긍정 지정사 개수
verb_count	동사 개수	pos_copula_count	부정 지정사 개수

<표 4> 메타 데이터 특성

특성 이름	설명	특성 이름	설명
dislike_percentage	영상의 싫어요 비율	words_count	영상 설명 단어 개수
view_count	영상 총 조회수	question_mark	영상 설명 물음표 유무
video_length_sec	영상 길이 (초)	exclamination_mark	영상 설명 느낌표 유무
view_per_day	일별 조회수	question_mark_count	영상 설명 물음표 개수
comments_per_day	일별 코멘트 개수	exclamination_mark count	영상 설명 느낌표 개수
text_length	영상 설명 텍스트 길이	donation_info	영상 설명 후원 계좌정보 유무

명사(일반명사, 고유명사), 형용사, 동사, 부사(일반부사, 접속부사), 지정사(긍정 지정사, 부정 지정사)의 출현 빈도를 특성으로 활용하였다(<표 3> 참조).

영상 이미지 특성은 마이크로소프트사의 Azure에서 제공하는 Face API를 사용하여 계산된 표정 점수를 추출하여 활용하였다. Face API는 총 8개 표정(분노, 경멸, 역겨움, 공포, 행복, 중립, 슬픔, 놀라움)에 대한 점수를 제공하며 개별 감정 점수의 합은 1이 된다. 대부분의 이미지에서 중립 감정에 대한 표정 점수가 높은 점을 고려하여 본 연구에서는 영상별로 중립 점수가 가장 낮은 이미지의 표정 점수를 사용하였다(<그림 2>의 step 6-2 참조).

메타 데이터 관련 특성은 기존 연구를 준용하여 (Vedova et al., 2018) 영상의 싫어요 비율, 영상 총 조회수, 영상 길이 (초), 일별 조회수, 일별 코멘트 개수, 영상 설명 텍스트 길이, 영상 설명 단어 개수, 영상 설명 물음표 유무, 영상 설명 느낌표 유무, 영상 설명 물음표 개수, 영상 설명 느낌표 개수, 영상 설명 후원 계좌정보 유무를 고려하였다(<그림 2>의 step 6-3 참조). 이를 요약하면 <표 4>와 같다.

본 연구의 목적은 가짜 뉴스 탐지를 위한 영상정보의 유용성을 파악하는 데 있다. 이를 위해서는 텍스트와 영상정보 각각의 특성뿐 아니라 이들의 다양한 조합을 비교 분석할 필요가 있다. 본 연구에서는 1) 텍스트 데이터 2) 메타 데이터 3) 표정 점수 4) 텍스트 데이터 + 메타 데이터 5) 텍스트 데이터 + 표정 점수 6) 메타 데이터 + 표정 점수 7) 텍스트 데이터 + 메타 데이터 + 표정 점수의 7가지 조합을 활용하여 학습과 평가를 위한 입력 값으로 사용하였다(<그림 2>의 step 7 참조).

3.3 데이터 분석

데이터 분석은 다양한 머신러닝과 딥러닝 기법을 활용한 모델 학습과 학습된 모델의 평가

를 포함한 2단계로 구성된다. 학습 모형은 특성의 효과 비교에 초점을 두기 위해 비교적 단순한 모형인 로지스틱 회귀분석(logistic regression), 나이브 베이즈(naïve bayes), 의사결정 나무(decision tree), 랜덤 포레스트(random forest), 서포트 벡터 머신(support vector machine)을 사용하였다 (<그림 2>의 step 8 참조). 각 분류 모형은 학습 데이터를 기반으로 10회 교차 검증을 수행하였으며 그리드 서치(grid search)를 통해 최적의 하이퍼파라미터(hyperparameter)를 선택하였다. 또한, 최신 딥러닝 기법을 활용하기 위해 딥뉴럴 네트워크(DNN: deep neural network) 모형을 사용하여 학습을 진행하였다.

학습을 통해 구축된 특성 조합별 분류 모형의 성능 평가를 위해 곡선아래면적(AUC: area under the curve)을 사용하였다. 곡선아래면적은 클래스가 불균형한 경우 정확도(accuracy)보다 견고성(robustness)이 높기 때문에 이를 주요 평가 기준으로 활용하였다. 다만 많은 인공지능 연구에서 모델 간 성능 평가를 위해 정확도를 사용하고 있기 때문에 이를 곡선아래면적과 함께 모델 간 성능 비교를 위한 보조 수단으로 활용하였다.

Ⅳ. 분석 및 논의

4.1 결과분석

본 연구에서 수집된 특성 조합의 기술통계를 요약하면 다음 <표 5>와 같다. <표 5a>에서 알 수 있듯이 409개의 영상에 대해 명사가 가장 많은(평균=354.645) 반면 부정 지정사가 가장 적은(평균 3.885) 것으로 나타났다. 가장 많이 나타난 감정은 중립(평균=0.871)인 반면 분노(평균=0.001), 역겨움(평균=0.001), 공포(평균=0.001)의 감정이 가장 적게 나타났다(<표 5b> 참조). 이는 객관적인 사실의 전달이라는 뉴스의 특징을 잘 나타낸다고 할 수 있다.

<표 5> 기술 통계량

<표 5a> 텍스트 변수 기술 통계량

	명사	형용사	동사	부사	부정 지정사	긍정 지정사
평균	354.645	24.932	115.276	59.756	3.885	21.169
표준편차	404.761	36.952	162.445	92.388	5.677	25.72
최소	24	0	3	1	0	0
최대	3322	270	1257	709	38	239

<표 5b> 표정 점수 기술 통계량

	분노	경멸	역겨움	공포	행복	중립	슬픔	놀라움
평균	0.001	0.007	0.001	0.001	0.041	0.871	0.027	0.045
표준편차	0.008	0.019	0.004	0.004	0.1	0.146	0.055	0.064
최소	0	0	0	0	0	0	0	0
최대	0.136	0.233	0.068	0.068	0.832	1	0.364	0.338

<표 5c> 영상 메타 데이터 기술 통계량

	싫어요 비율	총 조회수	일별 조회수	일별 조회수	일별 코멘트 수	설명 텍스트 길이
평균	0.187	25592.421	3089.756	1.578	0.476	704.389
표준편차	0.227	103762.33	18565.162	3.013	1.397	606.817
최소	0	9	30	0.18	0	0
최대	1	1598496	157200	30.77	14.85	4837

	설명 단어수	물음표 유무	느낌표 유무	물음표 개수	느낌표 개수	후원 계좌정보
평균	134.518	0.609	0.13	2.941	0.308	0.061
표준편차	138.774	0.489	0.336	11.81	1.32	0.24
최소	0	0	0	0	0	0
최대	1125	1	1	202	17	1

<표 6> 실험 설계

특성 조합	특성 조합 상세	분류 모형
개별 조합	(1) 텍스트	Logistic Regression Naïve Bayes Decision Tree Random Forest Support Vector Machine(SVM) Deep Neural Network(DNN)
	(2) 메타 데이터	
	(3) 표정 점수	
2개 조합	(4) 텍스트+메타 데이터	
	(5) 텍스트+표정 점수	
	(6) 메타 데이터+표정 점수	
3개 조합	(7) 텍스트+메타 데이터+표정 점수	

마지막으로 영상 메타 데이터의 경우 뉴스 영상의 평균 총 조회수는 약 25,500회였으며 일별 평균 조회수는 약 1.5회로 나타났다. 영상의 평균 길이는 약 3,000초이며 설명 텍스트의 평균 길이는 약 21로 나타났다(<표 5c> 참조).

본 연구에서는 7가지 특성 조합의 분류 성능을 파악하기 위해 머신러닝과 딥러닝 모형을 포함한 6가지 모형에 적용하여 42가지 실험을 진행하였다 (<표 6> 참조).

7개 특성 조합을 고려하여 각 분류 모형을 테스트 데이터 셋에 적용한 곡선아래면적은 <표 7>과 같다. 곡선아래면적은 분류 모델 성능을 측정한 그래프인 ROC(receiver operating characteristic) 곡선의 아래 면적을 의미하며, ROC는 x축을 특이도(specificity), y축을 민감도(sensitivity)로 하여 시각화한 그래프를 말한다. 특이도는 $\frac{TN}{TN+FP}$ 라는 식을, 민감도는 $\frac{TP}{TP+FP}$ 라는 식을 활용하여 각각 산출할 수 있다[4].

4) TN은 실제 거짓을 거짓으로 예측한 수를, TP은 실제 참을 참으로 예측한 수를, FP는 실제 거짓을 참으로 예측한 수를, FN은 실제 참을 거짓으로 예측한 수를 의미한다.

<표 7> 곡선아래면적을 활용한 특성 조합별 분류 성능

특성 조합* / 분류 모형	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Logistic Regression	0.516	0.602	0.519	0.648	0.545	0.659	0.659
Naïve Bayes	0.653	0.500	0.503	0.757	0.644	0.652	0.762
Decision Tree	0.600	0.706	0.497	0.706	0.509	0.675	0.675
Random Forest	0.669	0.655	0.530	0.681	0.676	0.644	0.624
SVM(support vector machine)	0.547	0.516	0.503	0.679	0.603	0.556	0.647
DNN(deep neural network)	0.717	0.710	0.691	0.746	0.737	0.741	0.752

* (1): 텍스트, (2): 메타 데이터, (3): 표정 점수, (4): 텍스트+메타 데이터, (5): 텍스트+표정 점수, (6): 메타 데이터+표정 점수, (7): 텍스트+메타 데이터+표정 점수

먼저 텍스트, 영상 메타 데이터, 영상 표정 점수를 각각을 개별적으로 사용한 모형의 경우 4개 모형(나이브 베이즈(naive bayes), 랜덤 포레스트(random forest), 서포트 벡터 머신(SVM), 딥 뉴럴 네트워크(DNN))에서 특성 조합 (1) 텍스트 특성을 사용했을 때의 곡선아래면적이 상대적으로 높은 것으로 나타났다. 다음으로 두 개 특성을 결합한 경우를 살펴보면 특성 조합 (4) 텍스트와 영상 메타 데이터를 결합했을 때가 의사결정 나무를 제외한 5개 모형에서 개별 특성을 사용한 경우보다 성능이 개선되는 것을 확인할 수 있다[5]. 로지스틱 회귀분석(logistic regression) 모형은 최대 13.2%p(0.648-0.516), 나이브 베이즈 모형은 최대 25.7%p (0.757-0.500), 랜덤 포레스트 모형은 최대 15.1%p(0.681-0.530), 서포트 벡터 머신 모형은 최대 17.6%p(0.679-0.503), 딥 뉴럴 네트워크 모형은 최대 5.5%p(0.746-0.691)의 성능 향상을 보였다. 특성 조합 (5) 텍스트와 영상 표정 점수를 결합한 경우 랜덤 포레스트, 서포트 벡터 머신, 딥 뉴럴 네트워크 모형에서 개별 특성을 사용한 경우에 비해 성능이 작게는 0.7%p (0.676-0.669)부터 크게는 14.6%p(0.676-0.530)까지 곡선아래면적이 증가하였다. 특성 조합 (6) 영상 메타 데이터와 영상 표정 점수를 조합한 경우에는 로지스틱 회귀분석, 서포트 벡터 머신, 딥 뉴럴 네트워크 모형에서 개별 특성을 사용한 경우보다 성능이 증가하였다. 마지막으로 모든 특성을 사용한 조합 (7)의 경우 로지스틱 회귀분석 모형이 0.659, 나이브 베이즈 모형이 0.762, 딥 뉴럴 네트워크 모형이 0.752로 개별 혹은 2개 특성을 조합한 경우보다 뛰어난 성능을 보였다. 이러한 결과는 영상정보(영상 메타 데이터, 영상 표정 점수)를 텍스트 특성과 함께 활용함으로써 가짜 뉴스 탐지 성능을 향상시킬 수 있음을 의미한다.

각 모형 별로 가장 높은 곡선아래면적 값을 확인한 결과 앞서 서술한 바와 같이 로지스틱 회귀분석, 나이브 베이즈, 딥 뉴럴 네트워크 모형에서 세 가지 특성 모두를 조합했을 때 (즉,

5) 의사결정 나무의 경우 텍스트와 메타 데이터의 결합 조합과 메타 데이터만을 이용한 경우 곡선아래면적이 0.706으로 동일한 값을 보이는 것으로 나타났다.

특성 조합 (7)인 경우) 곡선아래면적 값이 가장 높음을 확인하였다. 반면에 의사결정 나무, 랜덤 포레스트, 서포트 벡터 머신 모형은 텍스트와 영상 메타 데이터를 결합한 특성 조합 (4)인 경우에 가장 높은 곡선아래면적 값을 보였다. 이러한 현상은 영상 표정 점수 데이터의 특성에 기인했을 것으로 판단된다. <표 4b>에서 알 수 있듯이 대부분의 표정 점수가 8가지 감정 가운데 중립 감정에 분포되어 있다. 따라서 중립 감정을 제외한 나머지 감정이 차지하는 비율이 매우 적어 특성 조합 (4)가 특성 조합 (7)보다 상대적으로 높은 성능을 보이는 것으로 추정할 수 있다.

다음으로 보조 평가 지표인 정확도를 기준으로 특성 조합에 따른 분류 모형의 성능을 비교하였다 (<표 8> 참조). 정확도는 모형이 얼마나 정확한지를 평가하는 척도이며 $\frac{TP+TN}{TP+TN+FP+FN}$라는 수식을 통해 산출할 수 있다.

분석결과 개별 모형의 경우 곡선아래면적을 사용한 분류 결과 유사하게 4개 모형 (나이브 베이즈, 랜덤 포레스트, 서포트 벡터 머신, 딥 뉴럴 네트워크)에서 특성 조합 (1) 텍스트 특성을 사용했을 때의 정확도가 상대적으로 높은 것으로 나타났다. 텍스트와 영상 메타 데이터를 결합한 특성 조합 (4)의 경우 곡선아래면적을 사용한 분류 결과와 유사하게 의사결정 나무와 서포트 벡터 머신을 제외한 4개 모형에서 개별 특성을 사용한 경우보다 성능이 개선되는 것을 확인할 수 있다[6]. 특성 조합 (5) 텍스트와 영상 표정 점수를 결합한 경우 역시 곡선아래면적을 사용한 분류 성능과 유사하게 나이브 베이즈, 랜덤 포레스트, 서포트 벡터 머신 모형에서 개별 특성을 사용한 경우에 비해 성능이 작게는 0.12%p(0.772-0.764)부터 크게는 0.73%p (0.772-0.699)까지 정확도가 증가하였다. 특성 조합 (6) 영상 메타 데이터와 영상 표정 점수를

<표 8> 정확도를 활용한 특성 조합별 분류 성능

분류 모형 \ 특성 조합*	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Logistic Regression	0.748	0.756	0.707	0.764	0.732	0.780	0.780
Naïve Bayes	0.756	0.740	0.699	0.805	0.772	0.740	0.813
Decision Tree	0.724	0.805	0.675	0.805	0.724	0.789	0.789
Random Forest	0.780	0.789	0.740	0.813	0.805	0.789	0.789
SVM(support vector machine)	0.764	0.748	0.699	0.764	0.772	0.748	0.748
DNN(deep neural net)	0.789	0.764	0.699	0.813	0.772	0.780	0.821

* (1): 텍스트, (2): 메타 데이터, (3): 표정 점수, (4): 텍스트+메타 데이터, (5): 텍스트+표정 점수, (6): 메타 데이터+표정 점수, (7): 텍스트+메타 데이터+표정 점수

6) 의사결정 나무의 경우 텍스트와 메타 데이터의 결합 조합과 메타 데이터만을 이용한 경우 정확도가 0.805로 동일한 값을 보였으며, 서포트 벡터 머신의 경우 텍스트만을 이용한 경우와 정확도가 0.789로 일치하였다.

조합한 경우 역시 곡선아래면적을 활용한 결과와 유사하게 로지스틱 회귀분석, 의사결절 나무, 랜덤 포레스트 모형에서 개별 특성을 사용한 경우보다 성능이 증가하였다. 마지막으로 모든 특성을 사용한 조합 (7)의 경우 로지스틱 회귀분석 모형이 0.780, 나이브 베이즈 모형이 0.813, 딥 뉴럴 네트워크 모형이 0.821로 개별 혹은 2개 특성을 조합한 경우보다 뛰어난 성능을 보였다. 이러한 결과 또한 영상정보(영상 메타 데이터, 영상 표정 점수)를 텍스트 특성과 함께 활용함으로써 가짜 뉴스 탐지 성능을 향상시킬 수 있음을 의미한다.

각 모형별로 가장 높은 정확도 값을 확인한 결과 앞서 서술한 바와 같이 로지스틱 회귀분석, 나이브 베이즈, 딥 뉴럴 네트워크 모형에서 세 가지 특성 모두를 조합했을 때(즉, 특성 조합 (7)인 경우) 정확도 값이 가장 높음을 확인하였다. 반면에 의사결정 나무와 랜덤 포레스트 모형은 텍스트와 영상 메타 데이터를 결합한 특성 조합 (4)가 서포트 벡터 머신 모형은 텍스트와 영상 표정 점수를 결합한 특성 조합 (5)가 가장 높은 정확도 값을 보였다. 이러한 결과 역시 앞서 언급한 바와 같이 영상 표정 점수 데이터의 특성에 기인했을 것으로 판단된다.

4.2 함의

본 연구는 다음과 같은 점에서 학문 및 실무적 의의가 있다. 먼저 학문적 측면에서 보면 본 연구는 첫째, 영상정보를 활용한 가짜 뉴스 탐지 방법을 제시함으로써 기존 가짜 뉴스 탐지 방식의 단점을 보완하였다는 점에서 의의가 있다. 가짜 뉴스 탐지와 관련한 기존 연구들은 텍스트 정보 중심으로 진행되어 왔다. 이러한 방식은 “뉴스 읽기”라는 전통적 뉴스 소비 방식과 매우 밀접하게 연관되어 있을 뿐만 아니라 감성, 문체적 특징과 같이 텍스트로부터 획득할 수 있는 다양한 정보를 체계적으로 고려함으로써 가짜 뉴스 탐지 연구에 일정 정도 기여하였다. 그러나 텍스트 중심의 가짜 뉴스 탐지 방식은 소셜 미디어의 급격한 발전으로 인해 텍스트와 이미지를 결합한 뉴스 기사의 증가와 가짜 뉴스의 주요 원천으로 이미지의 중요성이 증가로 인해 더 이상 실용적이지 않게 되었다(Guo et al. 2020). 이러한 한계점을 극복하고자 많은 연구들이 텍스트와 이미지 정보를 결합하여 가짜 뉴스 탐지를 시도하였으며 가짜 뉴스 탐지 성능을 일정 정도 개선한 것도 사실이다. 그러나 이러한 조합 방식 역시 “뉴스 보기”로 변화하고 있는 영상정보 중심의 최신 뉴스 소비 추세를 반영하지 못한다는 점에서 한계가 있다. 본 연구는 이러한 문제점을 해결할 수 있는 방안을 제시함으로써 가짜 뉴스 탐지 연구의 지평을 넓힐 수 있을 것으로 기대한다. 둘째, 뉴스를 진행하는 화자의 표정과 같은 직접적인 영상정보를 활용함으로써 기존 영상정보 기반 가짜 뉴스 탐지 연구를 확대 발전시켰다는 점에서 의의가 있다. 최근 영상정보 중심의 뉴스 소비 추세를 반영하기 위하여 영상정보 기반 가짜 뉴스 탐지 연구가 제한적이나마 진행되고 있다(Serrano et al., 2020). 그러나 이러한 연구들은 가짜 뉴스 탐지를 위해 활용할 수 있는 다양한 정보 가운데 영상 메타 데이터의 극히 제한적인 부문만을 사용함으로써 진정한 의미의 영상정보 기반 가짜 뉴스 탐지 연구라 보기 어렵다. 본 연구에서는 영상 설명, 좋아요, 싫어

요, 제목, 업로드 날짜, 영상 길이, 조회수 등의 다양한 영상 메타 데이터를 활용하여 기존 연구의 단점을 극복할 수 있을 것으로 기대한다. 마지막으로 가짜 뉴스 탐지에 있어 텍스트 정보, 화자의 표정으로 표현되는 영상정보, 영상 관련 메타 데이터로부터 다양한 특성을 추출하고 이들 조합의 효과성을 검증하여 가짜 뉴스 탐지를 위한 가장 효과적인 조합을 파악함으로써 기존 연구를 확대 발전시켰다는 점에서 의의가 있다. 영상정보 중심 뉴스의 경우 영상 대본이라는 텍스트 정보, 뉴스 화자의 영상 이미지 정보, 뉴스 영상 메타 데이터 등 가짜 뉴스 탐지를 위한 다양한 특성 추출을 위한 소스(source)가 존재함에도 불구하고 기존 연구들은 주로 개별 정보에 기반하여 추출된 특성만을 활용하여 가짜 뉴스를 탐지하고 있다. 이로 인해 가짜 뉴스 탐지의 성능이 연구자에 따라 다르게 나타나고 결과의 일관성이 결여된 것도 사실이다. 본 연구에서는 다양한 특성 조합을 활용하여 가짜 뉴스 탐지를 시도함으로써 이러한 문제점을 어느 정도 해결할 수 있을 것으로 기대한다.

실무적 관점에서 보면 본 연구는 첫째, 기업이나 국가의 브랜드 이미지 또는 리스크 관리 담당자로 하여금 가짜 뉴스를 보다 정확하게 탐지할 수 있도록 지원함으로써 효과적인 브랜드 이미지 및 리스크 관리를 위한 가이드라인을 제시할 수 있을 것으로 기대된다. 특히 가짜 뉴스로 인한 피해가 점점 증가하고 있는 현실을 고려할 때 기업이나 국가의 가짜 뉴스에 대한 선제적 대응은 그 어느 때보다 중요하다 할 것이다. 둘째, 가짜 뉴스로 인한 방송 언론의 신뢰도 하락을 방지할 수 있는 실질적인 방안을 제시할 수 있을 것으로 기대된다. 영상 뉴스 서비스를 제공하고 있는 방송사들은 본 연구에서 제안한 방법론을 기존 팩트체크 서비스와 연계하여 활용함으로써 보다 효과적인 가짜 뉴스 탐지가 가능할 것이다. 마지막으로 유튜브와 같은 영상 서비스 제공자에게 가짜 뉴스를 빠르게 탐지할 수 있는 실질적인 방안을 제시함으로써 가짜 뉴스 확산을 조기에 방지할 수 있는 새로운 방안을 제시할 수 있을 것으로 기대된다. 특히 뉴스 화자의 영상정보를 기반으로 계량화된 표정 점수를 제공함으로써 보다 정교한 가짜 뉴스 관리가 가능할 것이다.

Ⅴ. 결 론

가짜 뉴스 탐지는 주요한 연구과제로 오랫동안 연구되고 있다. 특히, 소셜 미디어를 통한 뉴스 소비가 빠르게 증가함에 따라 가짜 뉴스로 인한 사회·경제적 문제는 더욱 중요한 연구과제가 되었다. 본 연구는 텍스트 위주의 기존 가짜 뉴스 탐지 연구의 한계점을 극복하기 위하여 텍스트 정보와 영상정보를 동시에 고려하여 가짜 뉴스를 탐지하고자 하였다. 이를 위해 뉴스를 진행하는 화자의 얼굴 표정을 점수화하고 이를 데이터로 입력하여 머신러닝과 딥러닝을 포함한 6개 모형에 적용 가짜 뉴스 분석을 시도하였다. 분석결과 텍스트, 영상 메타 데이터, 영상 표정 점수를 모두 조합한 방법이 대다수의 학습 모형에서 곡선아래면적과 정확도 측면에서 가장 우수한 것으로 나타났다. 이를 통해 텍스트 특성과 함께 영상정보를 사용함으로써 가짜 뉴스 탐지 모형의 성능 향상이 가능함을 확

인하였다. 본 연구는 영상정보를 활용하여 가짜 뉴스 탐지를 시도하였다는 점에서 학문적 의의가 있으며 가짜 뉴스로 인한 기업이나 국가의 피해를 예방할 수 있는 방안을 제시하였다는 점에서 실무적 의의가 있다.

그러나 본 연구는 다음과 같은 한계점이 있으며 이를 해결하기 위한 향후 연구가 필요하다. 첫째, 모형의 분석에 있어 충분한 양의 데이터를 활용하지 못하였다. 본 연구에서는 화자의 얼굴이 나오는 영상을 대상으로 데이터를 수집하고 분석을 수행하였다. 하지만 실제 유튜브 상에 업로드되는 뉴스 영상은 음성과 텍스트, 또는 이미지로만 구성된 경우가 다수 존재하였다. 이로 인해 데이터 수집 및 전처리 과정에서 상당수의 영상이 제외되었다. 나아가 영상의 해상도 또는 인터넷 구성 환경으로 인해 해당 영상의 식별이 불가능한 경우가 다수 존재하였다. 그럼에도 불구하고 보다 정교한 모형의 구축을 위해서는 대규모의 데이터가 필수적이다. 따라서 향후 연구에서는 다양한 방법을 통해 충분한 양의 데이터를 확보할 필요가 있다. 둘째, 텍스트 데이터의 품질에 한계가 존재하였다. 본 연구에서는 자동 생성된 자막을 통해 영상 대본 텍스트를 수집하였다. 그러나 이 경우 아직 인공지능의 발전이 충분하지 못한 한계점으로 인해 문장의 완성도가 낮아 정확한 문맥을 파악하기 어려운 경우가 다수 존재하였다. 이를 해결하기 위해 향후 연구에서는 딥러닝 기술을 사용한 STT(speech to text) 방법을 통해 영상 속 음성을 텍스트 형태로 변환함으로써 보다 정교한 텍스트 데이터를 확보할 필요성이 있다. 셋째, 본 연구의 경우 정치 이슈에 한정되었다는 단점이 존재한다. 그러나 가짜 뉴스의 경우 정치 이슈 이외에도 사회, 경제, 기술 등 다양한 분야에 걸쳐 나타나고 있는 실정이다. 따라서 향후 연구에서는 뉴스 대상을 크게 확대하여 분석의 신뢰성 및 일반화 가능성을 확보할 필요가 있다.

Acknowledgement

심사과정에서 두 분의 심사위원과 편집위원장이 제시한 많은 유용한 의견과 제안은 본 연구의 향상에 기여하였음. 본 연구는 2020년 추계 경영정보학회 학술대회 논문을 수정 및 보완하여 확장된 연구임.

참고문헌

민진영, 이애리, "'좋아요'와 '싫어요'같은 간접적 사회적 정보의 방향과 강도는 온라인 뉴스 콘텐츠 댓글의 숙의의 질과 어떤 관련이 있는가? 토픽 모델링을 이용한 토픽 다양성 분석," 정보시스템연구, 제30권, 제4호, 2021, pp. 303-327.

박아란, 이소은, "디지털 뉴스 리포트 2020," 한국언론진흥재단, 2020.

신동훈, 신우식, 김희웅, "작성자 언어적 특성 기반 가짜 리뷰 탐지 딥러닝 모델 개발," 정보시스템 연구, 제31권, 제4호, 2022, pp. 1-23.

정민, 백다미, "가짜 뉴스 (fake news)의 경제적 비용 추정과 시사점," 한국경제주평, 제736권, 2017, pp. 1-15.

정호선, “콘텐츠 기반 변수 추출 방법에 의거한 가짜 뉴스 분류,” 이화여자대학교 석사 학위논문, 2019.

최진호, 박영흠, “디지털 뉴스 리포트 2022,” 한국언론진흥재단, 2022.

한혜주, 이경미, “소비자의 소셜 미디어를 통한 정보공유 활동에 대한 연구,” 소비자학연구, 제25권, 제2호, 2014, pp. 21-44.

현윤진, 김남규. “뉴스와 소셜 데이터를 활용한 텍스트 기반 가짜 뉴스 탐지 방법론,” 한국전자거래학회지, 제23권, 제4호, 2018, pp. 19-39.

Ajao, O., Bhowmik, D., and Zargari, S., “Sentiment Aware Fake News Detection on Online Social Networks”, *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2019, pp. 2507-2511.

Bhutani, B., Rastogi, N., Sehgal, P., and Purwar, A., “Fake News Detection using Sentiment Analysis,” *Proceedings of Twelfth International Conference on Contemporary Computing (IC3)*, 2019, pp. 1-5.

Bondielli, A., and Marcelloni, F., “A Survey on Fake News and Rumour Detection Techniques,” *Information Sciences*, Vol. 497, 2019, pp. 38-55.

Cao, J., Guo, J., Li, X., Jin, Z., Guo, H., and Li, J., “Automatic Rumor Detection on Microblogs: A Survey,” *arXiv*, 2018, Arxiv Preprint Arxiv:1807.03505.

Capuano, N., Fenza, G., Loia, V., and Nota, F. D., “Content-Based Fake News Detection With Machine and Deep Learning: A Systematic Review,” *Neurocomputing*, Vol. 530, 2023, pp. 91-103.

Fridrich, J., and Kodovsky, J., “Rich Models for Steganalysis of Digital Images,” *IEEE Transactions on Information Forensics and Security*, Vol. 7, No. 3, 2012, pp. 868-882.

Gelfert, A., “Fake News: A Definition,” *Informal Logic*, Vol. 38, No. 1, 2018, pp. 84-117.

Guo, B., Ding, Y., Yao, L., Liang, Y., and Yu, Z., “The Future of False Information Detection on Social Media: New Perspectives and Trends,” *ACM Computing Surveys*, Vol. 53, No. 4, 2020, Article 68.

Horne, B. D., and Adali, S., “This Just in: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News,” *Proceedings of the 11th International AAAI Conference on Web and Social Media*, 2017.

Hu, X., Tang, J., Gao, H., and Liu, H., “Social Spammer Detection with Sentiment Information,” *Proceedings of the 2014 IEEE International Conference on Data Mining*, 2014, pp. 180-189.

Huh, M., Liu, A., Owens, A., and Efros, A. A., “Fighting Fake News: Image Splice

Detection via Learned Self-consistency," *Proceedings of the European Conference on Computer Vision*, 2018, pp. 101-117.

Ito, J., Song, J., Toda, H., Koike, Y., and Oyama, S., "Assessment of Tweet Credibility with LDA Features," *Proceedings of the 24th International Conference on World Wide Web*, 2015, pp. 953-958.

Jang, S. M., and Kim, J. K., "Third Person Effects of Fake News: Fake News Regulation and Media Literacy Interventions," *Computers in Human Behavior*, Vol. 80, 2018, pp. 295-302.

Jarrahi, A., and Safari, L., "Evaluating the Effectiveness of Publishers' Features in Fake News Detection on Social Media," *Multimedia Tools and Applications*, Vol. 82, 2023, pp. 2913-2939.

Jin, Z., Cao, J., Zhang, Y., Zhou, J., and Tian, Q., "Novel Visual and Statistical Image Features for Microblogs News Verification," *IEEE Transactions on Multimedia*, Vol. 19, No. 3, 2016, pp. 598-608.

Kwon, S., Cha, M., Jung, K., Chen, W., and Wang, Y., "Prominent Features of Rumor Propagation in Online Social Media," *IEEE 13th International Conference on Data Mining*, 2013, pp. 1103-1108.

Luvembe, A. M., Li, W., Li, S., Liu, F., Xu, G., "Dual Emotion based Fake News Detection: A Deep Attention-weight Update Approach," *Information Processing & Management*, Vol. 60, No. 4, 2023, 103354.

Masciari, E., Moscato, V., Picariello, A., and Sperlí, G., "Detecting Fake News by Image Analysis," *Proceedings of the 24th Symposium on International Database Engineering & Applications*, 2020, pp. 1-5.

Papadopoulou, O., Zampoglou, M., Papadopoulos, S., and Kompatsiaris, Y., "Web Video Verification using Contextual Cues," *Proceedings of the 2nd International Workshop on Multimedia Forensics and Security*, 2017, pp. 6-10.

Pariser, E., The Filter Bubble: What the Internet is Hiding from You, Penguin Press, New York, 2011.

Perez-Rosas, V., Kleinberg, B., Lefevre, A., and Mihalcea, R., "Automatic Detection of Fake News," *arXiv*, 2017, Arxiv Preprint Arxiv:1708.07104.

Potthast, M., Kiesel, J., Reinartz, K., Bevendorff, J., and Stein, B., "A Stylometric Inquiry into Hyperpartisan and Fake News," *arXiv*, 2017, Arxiv Preprint Arxiv: 1702.05638.

Serrano, J. C. M., Papakyriakopoulos, O., and Hegelich, S., "NLP-based Feature Extraction for the Detection of

COVID-19 Misinformation Videos on Youtube," *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL*, 2020.

Shearer, E., More than Eight-in-ten Americans Get News from Digital Devices, 2021, https://www.pewresearch.org/fact-tank/2021/01/12/more-than-eight-in-ten-americans-get-news-from-digital -devices.

Shu, K., Mahudeswaran, D., Wang, S., Lee, D., and Liu, H., "FakeNewsNet: A Data Repository with News Content, Social Context, and Spatiotemporal Information for Studying Fake News on Social Media," *Big Data*, Vol. 8, No. 3, 2020, pp. 171-188.

Silverman, C., "Facebook is Turning to Fact-checkers to Fight Fake News," *BuzzFeed.News*, 2016, https://www.buzzfeednews.com/article/craigsilverman/facebook-and-fact-checkers-fight-fake - news.

Singh, V. K., Ghosh, I., and Sonagara, D., "Detecting Fake News Stories via Multimodal Analysis," *Journal of the Association for Information Science and Technology*, Vol. 72, No. 1, 2021, pp. 3-17.

Vedova, M. L. D., Tacchini, E., Moret, S., Ballarin, G., DiPierro, M., and de Alfaro, L., "Automatic Online Fake News Detection Combining Content and Social Signals," *Proceedings of the 22nd Conference of Open Innovations Association*, 2018, pp. 272–279.

Vosoughi, S., Roy, D., and Aral, S., "The Spread of True and False News Online," *Science*, Vol. 359, No. 6380, 2018, pp. 1146-1151.

Zubiaga, A., Lialata, M., and Procter, R., "Learning Reporting Dynamics during Breaking News for Rumour Detection in Social Media", *arXiv*, 2016, Arxiv Preprint Arxiv:1610.07363.

장 윤 호 (Chang, Yoon Ho)

국민대학교 데이터사이언스 석사학위를 취득하였다. 현재 바이브컴퍼니 S.C.I연구소 연구원으로 재직 중이다. 주요 관심 분야는 데이터 사이언스, 딥러닝 등이며 경영 정보학회 학술대회에서 발표를 하였다.

최 병 구 (Choi, Byoung Gu)

KAIST 경영공학 석사 및 박사학위를 취득하였다. 현재 국민대학교 경영대학 AI빅데이터융합경영학과 교수로 재직 중이다. 주요 관심분야는 소셜미디어 어낼리틱스, 데이터사이언스, 디지털 비즈니스 등이다.

<Abstract>

Fake News Detection on Social Media using Video Information: Focused on YouTube

Chang, Yoon Ho · Choi, Byoung Gu

Purpose

The main purpose of this study is to improve fake news detection performance by using video information to overcome the limitations of extant text- and image-oriented studies that do not reflect the latest news consumption trend.

Design/methodology/approach

This study collected video clips and related information including news scripts, speakers' facial expression, and video metadata from YouTube to develop fake news detection model. Based on the collected data, seven combinations of related information (i.e. scripts, video metadata, facial expression, scripts and video metadata, scripts and facial expression, and scripts, video metadata, and facial expression) were used as an input for taining and evaluation. The input data was analyzed using six models such as support vector machine and deep neural network. The area under the curve(AUC) was used to evaluate the performance of classification model.

Findings

The results showed that the ACU and accuracy values of three features combination (scripts, video metadata, and facial expression) were the highest in logistic regression, naïve bayes, and deep neural network models. This result implied that the fake news detection could be improved by using video information(video metadata and facial expression). Sample size of this study was relatively small. The generalizablity of the results would be enhanced with a larger sample size.

Keyword: Fake News Detection, Video Information, Video Metadata, Facial Expression, Machine Learning, Deep Learning

* 이 논문은 2023년 5월 12일 접수, 2023년 5월 31일 1차 심사, 2023년 6월 12일 게재 확정되었습니다.