ISSN 2288-4866 (Print)
ISSN 2288-4882 (Online)
https://www.jiisonline.org
J Intell Inform Syst 2023 September: 29(3): 19~36 http://dx.doi.org/10.13088/jiis.2023.29.3.019

# 관련 동영상 정보를 활용한 YouTube 가짜뉴스 탐지 기법*

김준호
국민대학교 비즈니스IT전문대학원
*(junhokim@kookmin.ac.kr)*

신용준
강원대학교 컴퓨터공학과
*(sindydwns@kangwon.ac.kr)*

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

정보통신기술의 발전으로 인해 누구나 쉽게 정보를 생산, 유포할 수 있게 되면서, 이를 악용하여 의도적으로 유포하는 거짓 정보인 가짜뉴스가 새로운 문제로 대두되기 시작하였다. 초기에 텍스트 방식으로 주로 전파되던 가짜뉴스는 점차 진화하여 이제는 멀티미디어 형식으로 퍼지고 있다. 유튜브는 2005년에 설립된 이후 세계 최고의 동영상 플랫폼으로 성장하면서 전 세계 사람들이 대부분 이용하고 있다. 하지만 유튜브는 가짜뉴스가 퍼지는 주요 창구가 되며 사회적인 문제를 일으키고 있다. 유튜브의 가짜뉴스를 탐지하기 위하여 다양한 학자들이 연구를 진행해 왔다. 가짜뉴스 탐지 연구에는 콘텐츠 기반의 접근과 배경정보 기반의 접근이 존재하는데 기존 가짜뉴스 연구와 유튜브의 가짜뉴스 탐지 연구를 살펴보면 콘텐츠 기반의 접근이 다수를 차지하고 있다. 본 연구에서는 콘텐츠 기반의 가짜뉴스 탐지가 아닌 배경정보 기반의 가짜뉴스 탐지기법을 제안하는데, 그 중에서도 유튜브에서 제공하는 관련 동영상 정보를 활용하여 가짜뉴스를 탐지하는 방법을 제안하고자 한다. 구체적으로 관련 동영상에서 얻은 정보와 원본 동영상에서 얻은 정보를 임베딩 기술인 Doc2vec을 이용하여 벡터화 한 후, 딥러닝 네트워크인 합성곱 신경망(CNN)을 통하여 가짜뉴스를 판별하고자 하였다. 실증분석 결과 제안 기법은 기존의 콘텐츠 기반으로 유튜브 가짜뉴스를 탐지하는 접근에 비해 보다 우수한 예측 성능을 보임을 확인하였다. 이러한 본 연구의 제안 기법은 파급력이 높은 유튜브 상에서 유포되는 가짜뉴스의 전파를 사전에 예방함으로써, 우리 사회를 보다 안전하고 신뢰할 수 있도록 만드는데 기여할 수 있을 것으로 기대한다.

**주제어** : 가짜뉴스 탐지, 유튜브, Doc2vec, 합성곱 신경망, 관련 동영상

논문접수일 : 2023년 5월 15일 논문수정일 : 2023년 5월 15일 게재확정일 : 2023년 5월 30일
원고유형 : 학술대회 우수논문 교신저자 : 안현철

## 1. 서론

4차 산업혁명 시대를 살아가고 있는 우리는 전 세계 어디서나 접속할 수 있는 인터넷을 통해 서로가 연결되어 있고, 이를 통하여 정보를 수집하고 있다. 인터넷 신문을 포함한 수많은 웹사이트에서 다양한 정보에 누구나 접근할 수 있게 되었고 유튜브(YouTube)를 통해 다양한 동영상을 서로 공유하며 멀티미디어를 공유하게 되는 시대가 되었다. 과거에는 한정적인 매체로만 정보를 얻을 수 있어 정보의 양이 많지는 않았지만 정보의 신뢰도가 상당히 높았다(이원상, 2019). 국가 혹은 언론사에서 진위여부를 명확히 가려낸 이후 정보를 내보냈기 때문에 정보의 질적인 측면은 보장되었다. 하지만 현재는 인터넷 상의 수많은 정보를 누구나 쉽게 생성하고 전파할 수 있게 되면서 정보의 진위여부가 확실하지 않은 정보들이 상당히 많아 정보의 질적인 측면은 과거

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

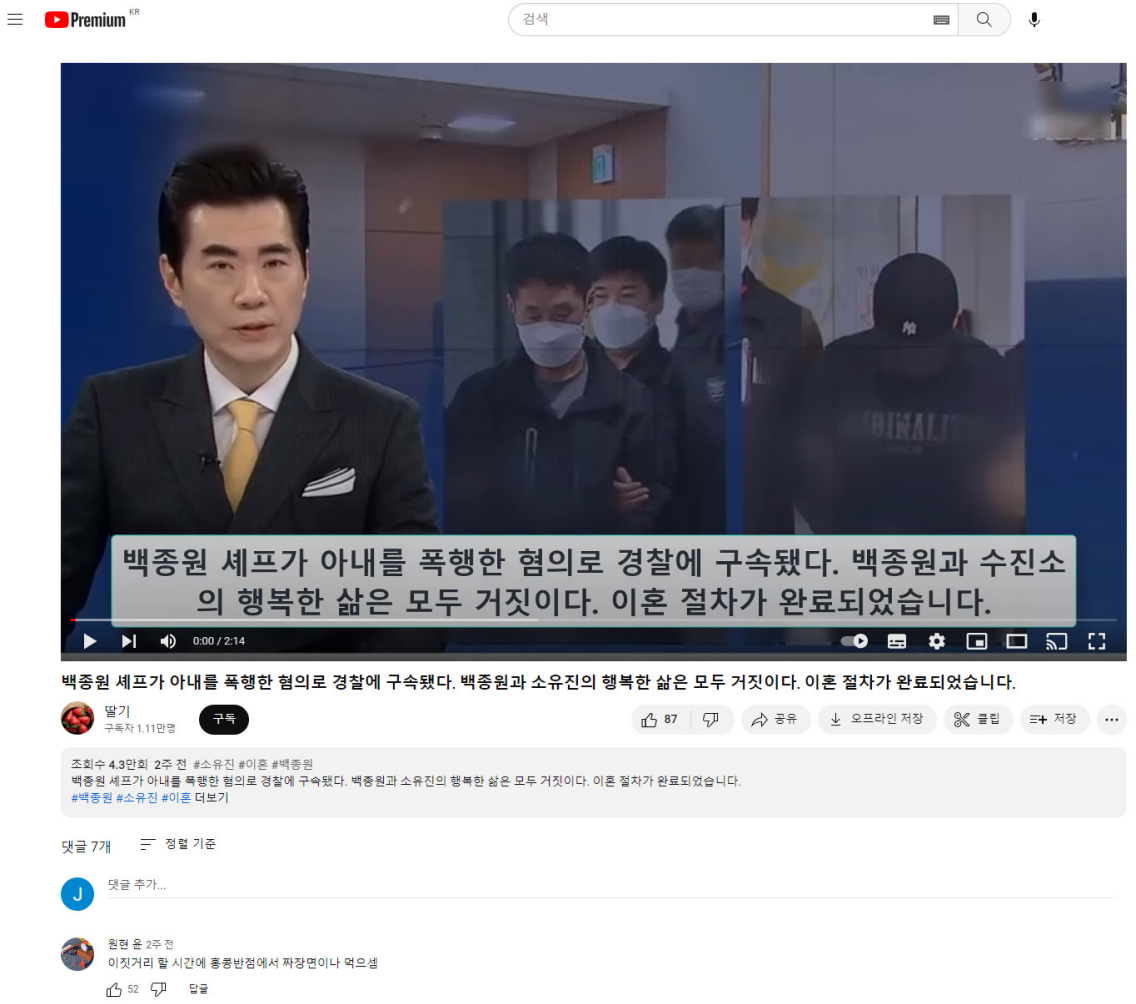
보다 하락하여 풍요 속의 빈곤인 실정이다. 또한 누구보다도 진위여부를 확인해야 할 기자들 역시 진위여부를 확인하지 않고 속보에 중점을 둬 언론사의 여과기능이 유명무실해지고 정보의 질적 측면에서의 하락을 불러왔다(김유나, 2021). 과거에도 항상 진위여부가 명확한 정보만 나온 것만은 아니다. 소위 증권사 찌라시와 같은 진위여부가 가려지지 않은 정보가 유통되곤 했지만 한정된 사람들만 접근할 수 있어 사회적 파장 자체는 적거나 미미하였다. 하지만 현재는 진위여부가 가려지지 않은 정보가 불특정 다수에게 빠른 속도로 유포되고 있으며, 이로 인한 사회적 파장은 과거와는 비교할 수 없을 정도로 폭발적이고 광범위하다(이원상, 2019).

일반적으로 가짜뉴스는 이처럼 진위여부가 가려지지 않은 정보를 마치 진짜인 것처럼 이익을 목적으로 의도적으로 뉴스의 형식을 가지고 유포하는 것을 의미한다(황용석, 권오성, 2017; 염정윤, 정세훈, 2019). 처음에는 단순히 텍스트로 이루어진 가짜뉴스가 주를 이루었다. 하지만 가짜뉴스도 점점 진화하였고 이제는 사진, 동영상 등 다양한 멀티미디어 형식을 부가하여 구체적이고 생생한 특징을 가진 형태로 진화해 가고 있다(좌희정 등, 2019). <그림 1>에 예시된 유튜브 상에서 유포되는 가짜뉴스의 사례에서 볼 수 있듯이, 멀티미디어 형태의 가짜뉴스는 정보수용자들의 신뢰성을 높이고, 오감을 자극해 과거 텍스트 형식보다 사람들의 뇌리에 보다 확실하게 각인하는 효과를 가져왔다. 그래서 멀티미디어 방식의 가짜뉴스를 판별하고 분류한 이후 가짜뉴스의 재생산을 막는 것은 사회적 건강도를 증진하기 위해 대단히 중요하다.

유튜브는 2005년에 설립된 이후 구글(Google)에 인수되어 현재는 전 세계에서 가장 많은 사람들이 찾는 동영상 플랫폼이 되었다. 전 세계적으로 유튜브에 로그인을 하는 사용자가 20억 명이 넘고, 영상의 재생시간도 하루에 10억 시간이 넘는다. 또한 특정 연령층이나 성별에 집중된 것이 아니라 남녀노소 가리지 않고 전부 유튜브를 사용한다. 통신 인프라가 우수한 한국에서는 유튜브의 영향력이 특히 더 강하다. 지난 2022년 한국인이 가장 많이 사용한 애플리케이션이 유튜브였고, 총 사용시간은 무려 175억시간에 달한다(임혜선, 2023). 이처럼 유튜브는 전 세계 많은 사람들의 정보수용처가 되었다. 이러한 유튜브에 <그림 1>에 예시된 것과 같은 가짜뉴스가 업로드 된다면, 그 정보는 한국을 넘어 전 세계로 퍼지게 되고 가짜뉴스로 인한 피해자의 수도 상당히 많을 것이며 결과적으로 이로 인한 사회적 파장은 막대할 것이다.

가짜뉴스를 탐지하는 것은 사회적, 경제적 효용성이 크고 정보의 신뢰성 제고를 위해 매우 중요한 일이므로 다양한 학자들이 가짜뉴스를 탐지하기 위해 노력하였다. 현재 학자들은 자동화된 가짜 뉴스 탐지를 크게 두 가지 방법으로 접근하고 있다. 그 중 하나는 가짜뉴스의 콘텐츠(contents)를 중심으로 접근하는 방법이며, 다른 하나는 가짜뉴스의 배경정보(context)를 활용하는 방법이다. 이 중 현재 가짜뉴스 탐지 연구의 주류를 차지하고 있는 기존 연구들은 대체로 콘텐츠를 중심으로 접근하는 방법을 채택하였다(Choi & Ko, 2021; Choi & Ko, 2022; Pan et al., 2018; Sheikhi, 2021; 이동호 등, 2018 등). 하지만 최근에 소개되고 있는 가짜뉴스 탐지의 연구들은 이러한 콘텐츠 정보의 활용 외에 배경정보의 활용을 강조하고 있다(박성수, 이건창, 2019; Shim et al., 2021; Raza & Ding, 2022 등).

멀티미디어 시대가 도래하며 유튜브 상의 가짜




〈그림 1〉 유튜브 상의 가짜뉴스 예시

뉴스를 탐지하는 것이 중요해지면서, 최근 멀티미디어 가짜뉴스 탐지를 위한 연구들이 시도되고 있다. 예를 들어, 장윤호와 최병구(2020)는 유튜브 가짜뉴스 탐지를 위해 영상과 텍스트 정보를 결합하여 가짜뉴스를 탐지하는 기법을 제안하였고, Choi & Ko(2022)는 도메인 지식과 멀티모달 데이터 퓨전(multimodal data fusion) 기법을 사용하여 유튜브 상의 가짜뉴스 비디오를 탐지하는 기법을 제안하였다. 하지만 이러한 기존 유튜브 상의 가짜뉴스 탐지 연구들은 모두 콘텐츠 중심으로 접근하는 방법을 채택하고 있으며, 배경정보를 기반으로 접근한 연구는 지금까지 거의 찾아보기 어려웠다.

한편 유튜브와 같은 동영상 플랫폼에서는 각 동영상을 재생했을 때 관련 동영상의 리스트를 제공하는데, 이러한 정보도 가짜뉴스를 판별하는데

유용한 정보가 될 수 있다. 관련 동영상의 리스트는 본 영상과 밀접하게 연관된 내용을 자동으로 제공해 주는데, 이 내용이 가짜뉴스와 진짜뉴스에 따라 그 구성이 확연히 달라질 수 있기 때문이다. 만약 이러한 배경정보를 기존의 콘텐츠 기반 접근법에 결합하여 시너지를 나게 한다면, 더욱 효과적인 가짜뉴스 판별이 가능할 것임을 예상해 볼 수 있다.

이에 본 연구에서는 강력한 유튜브의 관련 동영상 추천 기능으로 인해 가짜뉴스를 재생하였을 때 이용자가 허위정보에 쉽게 반복 노출될 수 있는 위험을 지적하고 있는 기존 연구들(정정주 등, 2019; Flora & Juliana, 2019)의 우려를 역으로 이용하여, 관련 동영상 정보를 가짜뉴스 탐지의 새로운 정보원으로 활용하는 연구를 시도하였다. 구체적으로 본 연구에서는 진짜뉴스 혹은 가짜뉴스를 담은 유튜브 동영상의 제목(title), 해설(description), 댓글(comment)과 관련 동영상(related video)으로 추천된 영상들의 제목, 해설, 댓글을 함께 Doc2vec으로 학습시켜 벡터화하고, 이후 해당 벡터값을 합성곱 신경망(Convolutional Neural Network, 이하 CNN)에 적용하여 최종적으로 해당 동영상이 진짜인지 가짜인지 판별하게 하는 이분류 모델을 구축한다. 이어 제안된 기법의 성능을 검증하기 위해, 공개된 유튜브 가짜뉴스 데이터셋을 이용해 배경정보인 관련 동영상 정보를 추가로 활용했을 때와 활용하지 않았을 때 가짜뉴스 탐지 성능에 유의미한 차이가 있는지를 실증분석을 통해 확인해 보고자 하였다.

이후 본 논문의 구성은 다음과 같다. 2장에서는 논문에 사용된 다양한 이론과 기법들의 배경을 설명한다. 구체적으로 자동화된 가짜뉴스 탐지 및 유튜브 가짜뉴스 연구에 대한 기존 연구의 동향을 분석할 것이다. 이어 3장에서는 연구모형 및 연구모형에서 사용된 Doc2vec과 CNN에 대하여 상세히 설명한다. 4장에서는 실증 분석으로서 분석에 사용된 데이터셋과 비교모델의 정의, 연구모델을 구성하는 각각의 변수 등의 대해 설명하고 이를 통해 도출된 실험결과를 제시한다. 마지막 5장에서는 결론과 함께 연구의 시사점 및 한계점에 대해 설명한다.

## 2. 이론적 배경

### 2.1. 자동화된 가짜뉴스 탐지

가짜뉴스를 자동으로 탐지하는 기술은 사회적, 경제적 효용성이 크고 정보의 신뢰성 제고를 위해 매우 중요한 일이므로 많은 학자들이 관심을 갖고 연구하였다. 자동화된 가짜 뉴스 탐지는 크게 두 가지 방법으로 접근하고 있다 첫 번째 방향은 뉴스 자체의 콘텐츠를 활용하는 방법으로 가짜뉴스의 본질적인 내용을 토대로 탐지하는 방법이다. 두 번째 방향은 가짜뉴스의 배경정보를 활용하는 방법으로, 가짜뉴스의 내용이 아닌 주변 정보를 활용하여 가짜뉴스를 탐지하는 방법이다(Bondielli & Marcelloni, 2019).

먼저 콘텐츠 기반의 접근 방법을 활용해 가짜뉴스 탐지를 시도한 기존 연구들을 살펴보면, Buntain & Golebeck(2017)의 연구는 트위터에서 수집한 데이터를 분석하여 질문부호, 감탄부호, 인칭대명사, 웃는 이모티콘의 사용빈도 등을 토대로 가짜뉴스를 판별하였으며, Pan et al.(2018)은 지식 그래프(knowledge graph)를 생성하고, 이를 기반으로 B-TransE 모델과 TransE 모델을 사용하여 뉴스 기사의 엔티티 및 관계 임베딩을 생성한 후 이를 통해 가짜뉴스를 탐지하는 기법을 제안하였다. 한편 Sheikhi(2021)는 Whale Optimization Algorithm (WOA)와 Extreme Gradient Boosting

Tree(xgbTree) 알고리즘을 결합한 하이브리드 모델을 이용하여 가짜뉴스를 탐지하는 기법을 사용하였으며, Wynne & Wint(2019)는 TF-IDF와 문자 N-gram을 이용하고 여기에 Gradient Boosting 판별기를 적용하여 가짜뉴스를 탐지하였다. 국내 연구인 이동호 등(2018)은 한국어 가짜뉴스를 얕은 CNN 모델과 음절 단위로 학습된 단어 임베딩 모델인 Fasttext를 활용하여 탐지하는 방법을 제안하였다. 이러한 콘텐츠 중심의 접근법은 가짜뉴스의 원천이라고 할 수 있는 내용 자체로부터 해답을 찾는 방법이기 때문에, 다수의 연구자들이 가장 우선적으로 고려했던 접근법이었다. 하지만 가짜뉴스도 점차 진화하고 있어서, 콘텐츠 특성 상 진짜뉴스와 구별하기 어려운 보다 정교한 형태의 가짜뉴스들이 최근에 생성되고 있어 콘텐츠만으로 가짜뉴스를 판별하기가 점점 어려워지고 있다.

이러한 이유로 인해 최근 배경정보를 이용한 가짜뉴스 탐지 연구가 주목을 받고 있다. 전술했듯이 배경정보를 활용한 연구는 가짜뉴스의 내용을 보는 것이 아니라, 그 바깥을 보는 방법이다. 즉, 가짜뉴스가 있는 링크, 연관된 사이트, 연결된 소셜 네트워크 등 가짜뉴스의 콘텐츠 이외의 내용을 활용하는 방식이다. 관련 연구를 살펴보면, Shim et al.(2021)이 검색엔진인 Google의 검색결과 링크를 벡터화하는 Link2vec을 이용하여 가짜뉴스를 탐지하는 모델을 제안하였고, Raza & Ding(2022)은 콘텐츠와 소셜 배경정보를 이용하여 가짜뉴스를 탐지하는 방법을 제안하였다. 국내에서는 박성수와 이건창(2019)이 소셜미디어 뉴스 확산 네트워크에 주목하여 네트워크 임베딩 방법인 Deepwalk로 특징변수를 생성한 후, 여기에 로지스틱 회귀분석을 적용하여 가짜뉴스를 탐지한 바 있다.

## 2.2. YouTube 가짜뉴스

가짜뉴스의 시작은 단순한 텍스트로만 작성한 방식으로 출발하였다. 하지만 가짜뉴스도 설득력을 갖고, 정보 수용자의 신뢰를 높이기 위해 진화하였고 현재는 이미지 혹은 동영상을 추가하는 것과 같이 멀티미디어 형식의 가짜뉴스도 빠르게 확대되고 있다. 유튜브는 2005년 설립 이후 현재 전 세계 1위의 동영상 플랫폼이 되었다. 특히 한국에서는 모든 연령층에서 애플리케이션 사용시간 1위가 유튜브이며, 한국인 전체가 1년에 175억 시간 가량 사용한다고 한다(임혜선, 2023). 이러한 독점적 지위의 플랫폼인 유튜브에 가짜뉴스가 업로드 된다면 이는 전 세계의 정보 수용자에게 순식간에 전파되어, 사회적으로 큰 파장을 일으킬 수 있다. 이러한 사태의 심각성을 인지한 유튜브에서는 자체적으로 알고리즘을 통해 가짜뉴스를 삭제하고 있다고 자사 홈페이지에서 명시하고 있지만 유튜브 측에서 삭제한다고 해도 전부 찾아내서 삭제하는 것은 불가능하며 삭제되더라도 삭제된 가짜뉴스가 그대로 혹은 살짝 변경된 채로 지속적으로 업로드가 이루어지고 있고 순식간에 많은 사람들에게 공유가 이루어지고 있다. 이에 일부 학자들이 유튜브의 가짜뉴스를 빠르게 탐지할 수 있는 다양한 접근법을 제시하고 있으며, 이를 고도화하기 위해 노력하고 있다.

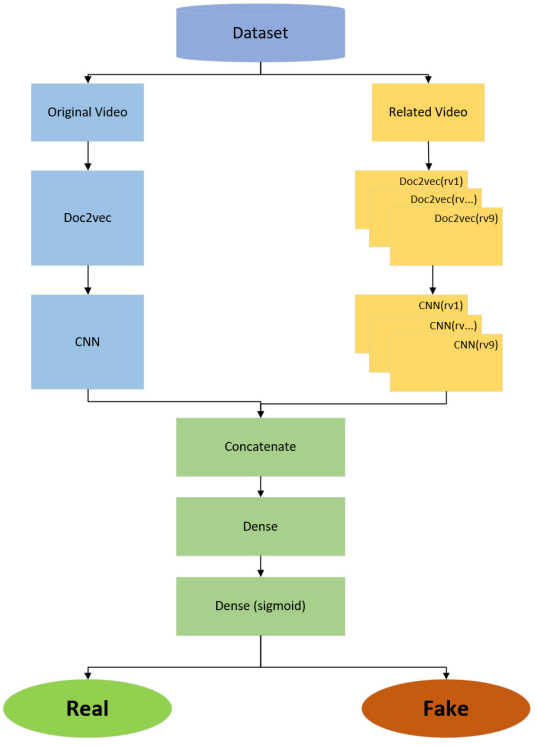
유튜브의 가짜뉴스를 탐지하는 연구는 다음과 같이 여러 방법으로 시도되어 왔다. 우선Yafooz et al.(2022)는 코로나19의 가짜뉴스를 아랍인과 아랍어의 특성에 맞춰 감정분석 및 딥러닝을 이용하여 탐지하였고, 장윤호와 최병구(2020)는 영상과 텍스트 정보의 결합한 이후 랜덤 포레스트 방식을 이용하여 가짜뉴스를 탐지하였다. 한편 Choi & Ko(2021; 2022)는 두 가지 방법을 제시하였는데,

우선 첫 번째로는 토픽 모델링과 적대적 인공신경망을 활용하여 가짜 뉴스를 탐지하는 기법을 제안하였고(Choi & Ko, 2021), 이어 두 번째로는 도메인 지식과 멀티모달 데이터 퓨전 기법을 사용하여 가짜뉴스 비디오를 탐지하는 방법을 제안하였다(Choi & Ko, 2022). 또한 Das et al.(2022)은 가짜뉴스 채널과 진짜뉴스 채널의 사용자에 대한 역학 모델링 후 모델에서 계산된 전달 계수 값을 사용하여 해당 채널의 시청자에서 구독자로의 사용자 이동을 추론하는 기법으로 가짜뉴스를 탐지하였다. 이상 소개한 기존 유튜브 가짜뉴스 탐지 연구는 대부분 콘텐츠 중심의 접근법을 적용하고 있으며, 아직 배경정보를 활용한 연구는 많지 않은 실정이다.

## 3. 연구모델

전술한 바와 같이 배경정보를 활용할 경우 콘텐츠만 단독으로 사용할 경우보다 가짜뉴스 탐지가 보다 효과적으로 수행될 수 있다는 최근의 연구들(Shim et al., 2021; Raza & Ding, 2022; 박성수, 이건창, 2019)이 존재함에도 불구하고, 유튜브 가짜뉴스 탐지를 주제로 한 기존 연구들은 주로 콘텐츠 중심의 접근법만을 사용하고 있다는 한계가 있다(Choi & Ko, 2021; 2022; Yagooz et al., 2022; 장윤호, 최병구, 2020). 이러한 한계를 극복하고자 본 연구에서는 유튜브가 제공하는 '관련 동영상(related videos)' 정보를 추가로 활용하여, 유튜브 가짜뉴스 탐지의 성능을 제고하는 새로운 접근법을 제안한다. 유튜브 가짜뉴스를 사회과학적 관점으로 접근한 기존 연구들(정정주 등, 2019; Flora & Juliana, 2019)에 따르면, 유튜브의 강력한 관련 동영상 추천 기능으로 인해 한 번 가짜뉴스에 노출된 이용자가 계속 유사한 정보를 전달하는 가짜뉴스에 반복 노출될 위험이 상당히 높아진다. 본 연구에서는 이와 같은 기존 연구의 우려를 역으로 이용하여, 유튜브가 추천하는 관련 동영상 정보를 가짜뉴스 판별의 새로운 배경정보원으로 활용하는 방법을 제안한다.

본 연구에서 제안하는 유튜브 가짜뉴스 탐지 모델의 구조는 다음의 <그림 2>와 같다.




〈그림 2〉 연구모델

수집된 데이터셋을 원본 동영상(original video)과 관련 동영상(related video)로 분리한 이후, 동영상의 제목, 설명, 댓글 등의 텍스트로 구성된 각 데이터셋 전체를 토큰화하고 Doc2vec을 학습하여 언어모델을 구축하였다. 먼저 원본 동영상은 Doc2vec을 이용하여 벡터화된 값을 추출한 이후,

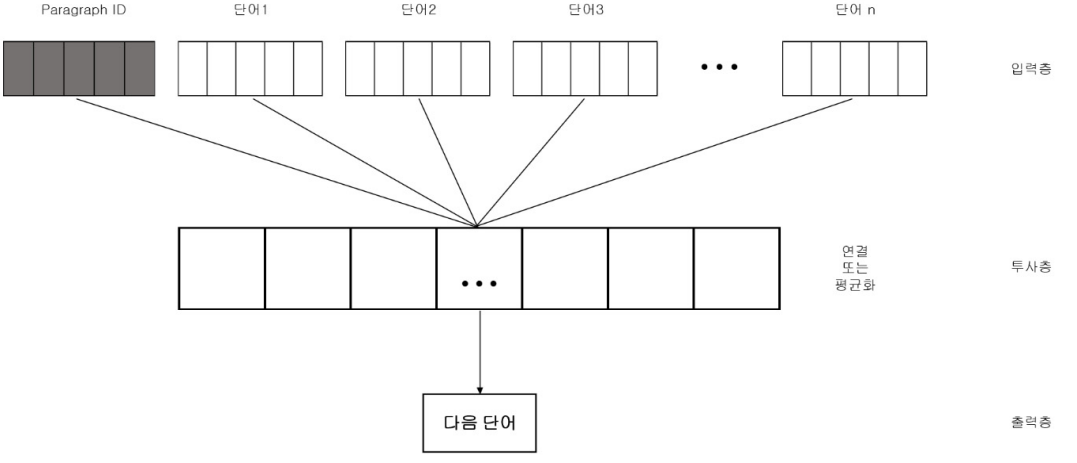
해당 벡터에 CNN을 적용하여 특징을 압축하였다. 그리고 관련 동영상은 1번부터 9번까지 총 9개 영상에서 수집된 텍스트 정보들을 각각 Doc2vec을 이용하여 벡터화된 값으로 변환하였고, 원본 동영상과 마찬가지로 이 변환된 값들에 다시 각각 CNN을 적용하여 특징을 압축하였다. 이렇게 압축, 정리된 특징들은 이후 하나로 연결(concatenate)되었으며, 이는 전연결층으로 연결되어 최종적으로 해당 원본 동영상이 가짜뉴스인지 혹은 진짜뉴스인지를 예측한다. 본 연구의 제안모델에서 주요하게 활용되고 있는 Doc2vec과 CNN에 대해 상세히 살펴보면 다음과 같다.

### 3.1. Doc2vec

Le & Mikolov(2014)가 발표한 Doc2vec은 Word2vec보다 한 단계 진화된 방식의 임베딩 방법이다. 기존의 Word2vec이 단어 간의 관계만 살펴봤다면 Doc2vec은 표현 그대로 문서 간의 관계를 살펴보는 모델을 기존의 Word2vec 방법과 결합한 방법이다. Word2vec이 각 문장의 문맥단위 단어를 학습하는 모델이라고 한다면, Doc2vec은 문맥에 등장하는 단어들의 분포적 특성을 추출하고, 이를 여러 문맥에서 응축하는 방식으로 활용한다. Doc2vec은 문장 전체에서 단어 k개가 주어지면, 다음 단어를 예측하는 과정을 반복하여 학습한다. 이 과정에서 평균 로그 확률을 최대로 만드는 방향으로 학습한다.

Doc2vec는 PV-DM 방식과 PV-DBOW 방식의 두 가지 방식이 있다. 이 중, PV-DM(Distributed Memory Model of Paragraph Vectors)은 <그림 3>과 같은 방식으로 단어 벡터들과 문맥 벡터를 연결하거나 평균화를 하여 다음 단어를 예측하는 방식이다. “나는 항상 아침마다 운동을 하고 있습니다.”라는 문장이 있을 때 창(window)이 2개라면, “나는”, “항상”의 단어 벡터들과 문맥 벡터에 해당하는 Paragraph ID를 통해 “아침마다”라는 다음 단어를 예측하는 것이다. 이 때 문서1과 문서2의 같은 단어, 예를 들어 “항상”이라는 단어를 통해 생성한 단어 벡터는 무조건 동일한 값을 가져야 한다.

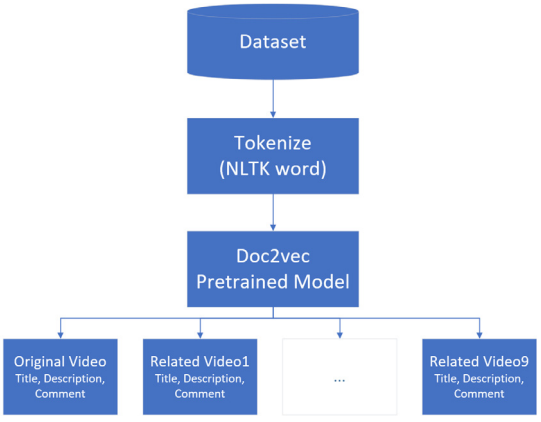
반면 PV-DBOW 방식은 Paragraph ID만 입력




〈그림 3〉 Doc2vec의 PV-DM 방식

하여 일정 개수의 단어를 예측하는 방법이다. 단, 단어들이 각각의 순서와 상관없이 랜덤으로 추출된다. 예를 들어 “나는 항상 아침마다 운동을 하고 있습니다.” 문장에서 단어를 3개 추출한다면, 앞선 순서대로 “나는”, “항상”, “아침마다”가 나오는 것이 아니고, “나는”, “운동을”, “항상”과 같이 랜덤 형식으로 단어가 배출되게 된다.

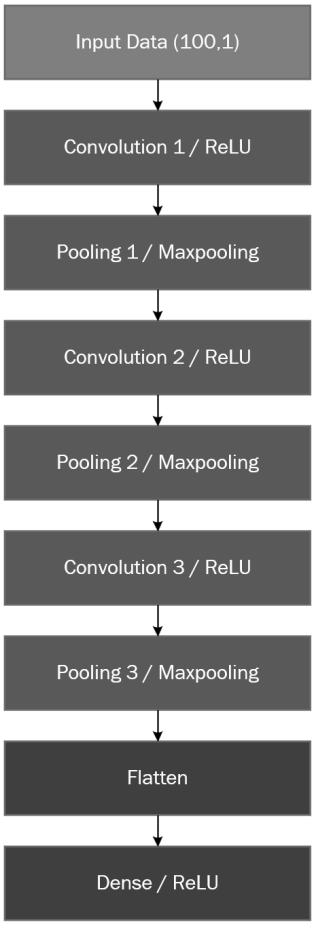
본 연구에서는 전술한 Doc2vec의 두 가지 구현방식 중에서는 PV-DM 방식을 적용한다. 아울러 본 연구에서는 <그림 4>와 같은 방식으로 모든 데이터셋을 이용하여 토큰 처리를 거쳐 미리 Doc2vec 언어모델을 사전 학습한 이후, 각각의 실험에 활용할 데이터를 언어모델을 이용해 벡터화 한 뒤 사용하였다.




〈그림 4〉 Doc2vec Pretrained Model 구축방법

### 3.2. CNN(Convolutional Neural Network)

CNN은 딥러닝 방법론 중 하나로 NYU의 Yann LeCun 교수가 발명한 합성곱(convolution)이라는 전처리 작업이 들어가는 인공신경망이다(LeCun et al., 2015). 본 연구에서는 Doc2vec으로 처리되어 100,1로 벡터화된 데이터를 <그림 5>와 같은 CNN 모델에 입력하였고, 이렇게 도출된 결과들을 이후 연결(concatenate)하여 활용하였다.




〈그림 5〉 본 연구에서 적용된 CNN의 구조

<그림 5>에서 데이터가 들어오면 합성곱 연산을 거치게 된다. 합성곱 연산이란 커널 또는 필터라 불리는 $n \times m$ 크기의 행렬로 높이 × 너비 크기의 이미지를 처음부터 끝까지 겹치게 훑으면서 $n \times m$ 크기의 겹쳐지는 부분의 각각의 이미지와 커널의 원소의 값을 곱한 후 모두 더한 값을 출력하는 것이다. 합성곱 연산 이후에는 풀링 층을

통과한다. 일반적으로 가장 많이 사용되면서 본 연구에서도 적용된 방법은 맥스 풀링(maxpooling)이다. 맥스 풀링이란 각 합성곱 연산으로부터 얻은 결과 벡터에서 가장 큰 값을 가진 스칼라 값을 추출하는 연산이다. 이러한 합성곱 연산과 풀링층은 CNN 모델 내에서 반복되는데, 본 연구에서는 총 3회 반복 실행하여 결과값을 추출하였다.

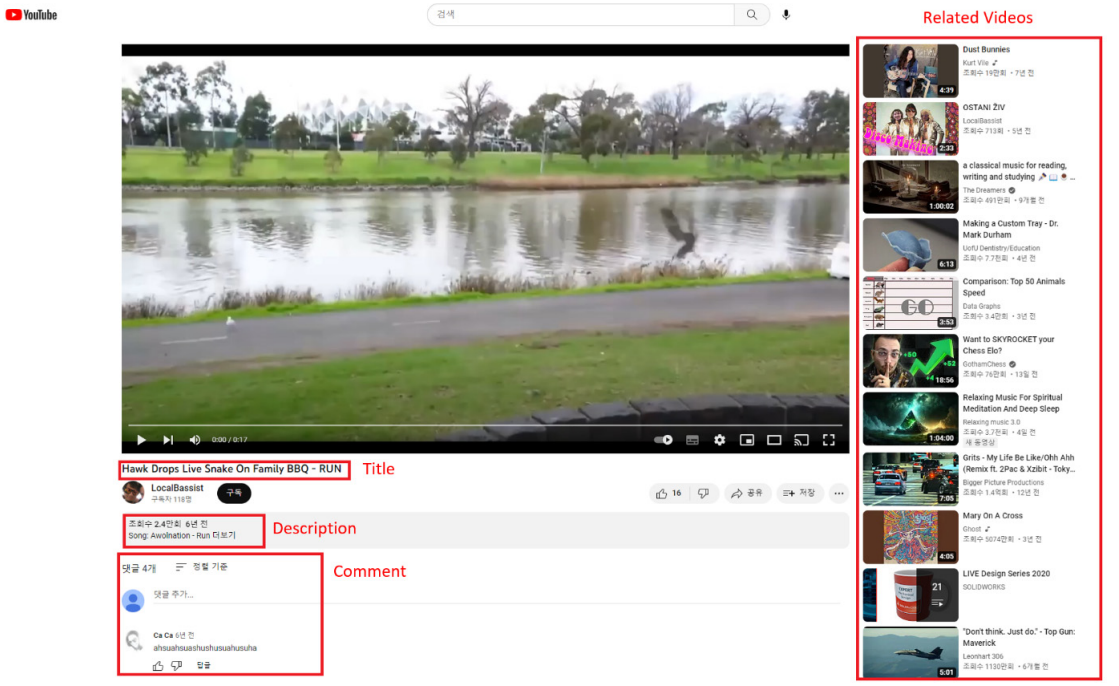
일반적으로 CNN은 이미지나 영상 데이터를 처리할 때 적용된다. 이미지의 공간적, 지역적 정보를 유지한 채 특성들의 계층을 형성하여 이미지의 전체보다는 각 부분을 보고 이미지의 픽셀과 주변 픽셀들의 연관성을 살리는데 유리하기 때문이다. 하지만, 본 연구에서는 이미지 대신 Doc2vec에서 받은 텍스트 데이터의 벡터값을 CNN의 입력으로 사용하였다. 그래서 이미지에서 사용한 2D 필터를 사용하지 않았고, 1D 필터를 사용하였다. 이미지를 처리할 때는 공간 관계를 고려하지만 텍스트를 처리할 때는 공간관계를 고려할 필요가 없기 때문이다. 임베딩된 단어를 처리할 때는 한 번에 단어 전체의 벡터를 고려하는 방식으로 이루어진다. 그래서 1D 필터를 사용하지만, CNN의 원리는 그대로 사용하는 네트워크를 설계하였다. 이렇게 CNN을 텍스트나 벡터에 활용하는 방식은 감정 분석, 질문유형 분석 등 다양한 분석에서 뛰어난 성능을 보여주고 있어 많은 연구자들이 사용하고 있으며 따라서 본 연구에서 사용하기 적합한 방식으로 판단된다.

## 4. 실증 분석

### 4.1. 실험 데이터

실험에 사용한 유튜브 영상은 MKLab의 공개 데이터셋인 Fake-Video-Corpus(FVC)이다(Papadopoulou et al., 2018). 본 데이터셋은 가짜뉴스, 진짜뉴스, 패러디 등 다양한 데이터를 포함하고 있는 데이터셋이며, 우리는 그 중에서 진위여부가 확실하게 구분된 진짜뉴스와 가짜뉴스의 데이터만을 사용하여 실험 데이터를 수집하였다. 본 데이터셋의 전체 비디오 숫자는 총 3028개였다. 이 중 패러디 등을 제외하면, 진짜뉴스는 1,015개, 가짜뉴스는 1,679개로 총 2,694개의 뉴스 데이터가 존재한다. 하지만 관련 동영상 데이터를 추가로 수집하는 과정에서 원본 동영상이 삭제되거나 비공개 처리된 경우가 다수 존재하여, 최종적으로는 진짜뉴스 654개, 가짜뉴스 1,018개의 데이터로 구성된 총 1,672개의 데이터를 실험에 사용하였다.

본 연구에서 API를 통해 수집한 유튜브 동영상의 데이터, 즉 제목(title), 설명(description), 댓글(comment) 및 관련 동영상(related videos)이 어떻게 실제 사용자에게 노출되는지에 대한 설명이 <그림 6>의 캡쳐 화면과 함께 제시되어 있다. 유튜브는 관련 동영상을 내보낼 때 이용자의 IP 주소를 이용하여 이용자의 국가 혹은 위치를 파악한 이후 각 지역 특성에 맞게끔 서로 다른 내용을 제공한다. 가령 한국에서는 한국어로 된 영상을 우선으로 관련 동영상을 생성한다. 또한 로그인이 되어 있는 경우, 개인별 시청목록 등을 토대로 관련 동영상을 생성하여 보여준다. 하지만 본 연구에서는 영상 정보를 수집할 때 동일한 위치에서 수집하여야 하며, 개인별 맞춤이 제공되지 않아야 일관성 있는 관련 정보를 수집할 수 있다는 점을 고려하여, Google에서 제공하는 YouTube API를 이용하여 구글 서버에서 직접적으로 관련 동영상 정보를 추출하였다. 그래서 미국 주소 기반으로 영상을 추출하게 되었고, API에서 직접적으로 추출한 영상이기 때문에 개인 로그가 없는 상태




〈그림 6〉 유튜브 동영상 조회 화면 예시와 수집된 데이터 현황

의 관련 동영상 정보를 추출할 수 있었다.

실험 데이터는 Google의 API를 이용하여, 유튜브 각 영상의 제목, 설명, 댓글의 데이터 및 관련 동영상 9개의 주소를 추출하였고, 관련 동영상 각각에 대한 제목, 설명, 댓글 데이터를 수집하였다. 모든 데이터는 SQL을 이용하여 개인 서버에 저장하였고, 이후 간단한 전처리를 거쳐 UTF-8 인코딩을 적용하여 다국적 언어 및 유니코드를 최대한 원형 그대로 살릴 수 있도록 하였다. 그리하여 총 1,672행, 33열의 데이터를 수집하였으며, 이를 json 파일로 저장하였다. 실험에 필요한 데이터셋의 분할을 위해 sklearn 라이브러리의 Train_Test_Split을 이용하여 학습용 60%, 검증용 20%, 시험용 20%로 분할하였다. 그리고 모든 실험과정에서 동일한 데이터셋을 활용하기 위해 Random_State는 42로 고정하여 사용하였다.

### 4.2. 실험 설계

원본 동영상의 제목(Title), 설명(Description), 댓글(Comment) 데이터를 Title, Description, Comment, Title+Description, Title+Comment, Description+Comment, Title+Description+Comment의 총 7가지의 경우로 분류하였고, 각각의 경우에 대해 독립적으로 분석을 진행하였다. 원본 동영상의 실험결과는 비교모델 및 실험모델에서 모두 사용되었다. 다음으로 관련 동영상을 1번부터 9번까지 순서대로 분리한 이후 관련 동영상 역시 원본 동영상처럼 Title, Description, Comment, Title+Description, Title+Comment, Description+Comment, Title+Description+Comment의 총 7가지의 경우로 구분

하였고, 관련 동영상 1번부터 순차적으로 9번까지 하나씩 추가해가며 모델을 학습하였다.

제안모델에서 관련 동영상의 개수를 몇 개로 확정할 것인가를 결정하는 문제에서는 1번부터 10번까지 순차적으로 추가하였을 때, 9번까지 추가한 모델이 가장 성능지표가 우수하였다. 이에 9번까지 추가한 모델을 제안모델로 확정하여 사용하였다.

한편 Doc2vec을 사전 학습할 때 전체 글자수가 7천만 단어인 것을 감안, vector size = 100, window 15, alpha=0.025, min_alpha=0.025, min_count=1, dm=1, negative=5, seed=9999로 세부 하이퍼파라미터를 지정하였다. 이 설정을 바탕으로 데이터셋의 모든 데이터를 5 epoch 학습하여 언어모델을 생성하였다. 이후 원본 동영상의 제목, 설명, 댓글 데이터와 관련 동영상 9개의 제목, 설명, 댓글 데이터를 각각 언어모델을 적용하여 (100,1)의 벡터 값으로 출력하여 저장하였다.

CNN은 Doc2vec으로 학습한 (100,1) 벡터값의 데이터를 입력층(input layer)으로 지정한 이후 텍스트 처리에 적합한 1D 합성곱층(convolution layer)을 사용한 CNN을 적용하였다. 이어 각각의 데이터를 CNN의 과정을 거친 이후 전부 연결(concatenate)하여 처리된 값을 합쳐주는 과정을 진행하였다. 이후 전연결층(dense layer)를 통과하여 최종 출력노드에 연결되도록 모델을 설계하였는데, 이 때 과적합 방지를 위해 Dropout을 적용하였다. 그 외에도 Validation Accuracy 및 Loss를 활용하여 Loss가 추가적으로 계속 상승한다면 조기종료(early stopping)를 통해 5회 이상 검증용 데이터셋의 Loss가 정체되거나 상승하면 학습을 자동 중단하게 설계하였다. 정확도를 포함하여 성과지표는 Accuracy, Precision, Recall, F1-Score, AUC 총 5개를 사용하였다.

모든 실험은 파이썬 3.9.13버전에서 구현하였다. 토크나이징에는 NLTK 라이브러리를 사용하였고 이후 Gensim 라이브러리의 Doc2vec을 사용하여 언어모델 구축 및 벡터화에 사용하였다. 그리고 벡터값을 추출한 이후 Tensorflow.Keras 라이브러리를 이용하여 CNN의 모델을 구축 및 정의하였다. 모델 구축 이후 다양한 성과지표를 출력하기 위해 sklearn을 로딩하여 혼동 행렬(confusion matrix)을 출력하였고, 이후 각 성과지표는 혼동 행렬을 통해 계산하여 출력하였다. AUC는 ROC_AUC를 통해 추출하였다. 또한 numpy, pandas는 파이썬의 기본적인 파일 불러오기 및 연산을 위해 필요하여 필수적으로 사용하였다.

### 4.3. 실험 결과

다음의 <표 1>에서는 모델 각각에 어떠한 정보를 활용하였는지 설명하고 있다. 모델ID의O는 원본 동영상의 정보를 활용했음을 의미하며, OR은 원본 동영상의 정보 및 관련 동영상들의 정보를 모두 활용했음을 의미한다. O-T와 OR-T는 제목만 활용한 모델이며, O-D와 OR-D은 설명만 활용한 모델, O-C와 OR-C는 댓글만 활용한 모델을 의미한다. O-TD와 OR-TD는 제목과 설명을 동시에 활용한 모델을 뜻하며, O-TC와 OR-TC는 제목과 댓글을 동시에 사용한 모델이다. O-DC와 OR-DC는 설명과 댓글을 동시에 사용한 모델이다. 끝으로 O-TDC와 OR-TDC는 제목, 설명, 댓글의 모든 정보를 활용하여 모델을 구성한 것을 나타낸다.

비교모델로 설정된 O로 시작하는 7가지의 모델은 원본 동영상의 정보만 가지고 Doc2vec 및 CNN을 활용하여 정확도 및 다양한 측정지표를 산출하였으며, 제안모델인 OR로 시작하는 7가지의 모델은 원본 동영상의 정보와 함께 관련 동영상의 정보를 활용하여 Doc2vec 및 CNN을 활용

〈표 1〉 실험에 사용된 데이터 분류

| 모델ID | 모델 구분 | | | | | |
|---|---|---|---|---|---|---|
| | Original Video | | | Related Videos | | |
| | Title | Description | Comment | Title | Description | Comment |
| O-T | O | | | | | |
| OR-T | O | | | O | | |
| O-D | | O | | | | |
| OR-D | | O | | | O | |
| O-C | | | O | | | |
| OR-C | | | O | | | O |
| O-TD | O | O | | | | |
| OR-TD | O | O | | O | O | |
| O-TC | O | | O | | | |
| OR-TC | O | | O | O | | O |
| O-DC | | O | O | | | |
| OR-DC | | O | O | | O | O |
| O-TDC | O | O | O | | | |
| OR-TDC | O | O | O | O | O | O |

하여 정확도 및 다양한 측정지표를 산출하였다. 그렇게 산출된 전체적인 성과지표 결과는 다음의 <표 2>에 제시되어 있다.

전술했듯이 본 연구에서는 분류 모델 평가를 위해 Train Set Accuracy, Validation Set Accuracy를 통해 과적합을 예방하고, 실험 결과를 평가하는 지표로는 Test Set Accuracy, Recall, Precision, F1_Score, AUC를 사용했다. 여기서 Recall(재현율)은 실제 가짜뉴스를 모델이 가짜뉴스라고 예측한 비율을 의미하고 Precision(정밀도)은 모델이 가짜뉴스라고 예측한 것들 중 실제 가짜뉴스의 비율을 의미한다. F1_Score은 Recall과 Precision의 조화평균으로서, 높을수록 모델 성능이 좋다. AUC는 임계 값을 변화시키면서 분류 문제에 대한 성능을 측정하는 지표로, 일반적으로 ROC는 확률 곡선을 의미하며, AUC는 분리의 정도를 나타내고, ROC 곡선 아래에 있는 면적을 의미한다. AUC의 값이 클수록 모델이 클래스를 잘 분류한다는 의미로 해석할 수 있다.

<표 2>에 제시되어 있는 전반적인 추세로 볼 때 원본 동영상과 관련 동영상들의 정보 모두를 활용한 제안모델인 OR이 보다 우수한 성능을 보여주고 있다. 일부 모델에서 Recall 값이 떨어지는 경우가 있긴 하지만 그 이외에 모든 지표가 O 모델보다 좋으며, 특히 사용된 데이터의 종류가 다양할수록 더 높은 성과를 보이는 것으로 나타났다. R-TDC 모델의 경우, F1-Score는 0.9269이라는 수치를 나타냈는데, 이는 Choi & Ko(2021)의

〈표 2〉 비교모델 및 제안모델 성과지표 측정치(%)

| 모델 ID | Accuracy | Precision | Recall | F1-Score | ROC_AUC |
|---|---|---|---|---|---|
| O-T | 66.87 | 70.97 | 81.86 | 76.03 | 69.65 |
| OR-T | 73.73 | 81.91 | 75.81 | 78.74 | 81.77 |
| O-D | 69.85 | 70.50 | 91.16 | 79.51 | 71.22 |
| OR-D | 84.48 | 86.88 | 89.30 | 88.07 | 91.78 |
| O-C | 65.67 | 70.16 | 80.93 | 75.16 | 63.04 |
| OR-C | 89.55 | 90.18 | 93.95 | 92.03 | 95.63 |
| O-TD | 68.96 | 77.07 | 73.49 | 75.24 | 74.47 |
| OR-TD | 84.78 | 85.34 | 92.09 | 88.59 | 92.34 |
| O-TC | 74.03 | 76.89 | 85.12 | 80.79 | 78.19 |
| OR-TC | 87.16 | 85.54 | 96.28 | 90.59 | 94.31 |
| O-DC | 77.31 | 78.84 | 88.37 | 83.33 | 77.86 |
| OR-DC | 88.36 | 94.44 | 86.98 | 90.56 | 95.88 |
| O-TDC | 70.75 | 77.99 | 75.81 | 76.89 | 78.57 |
| OR-TDC | 90.45 | 91.03 | 94.42 | 92.69 | 95.91 |

〈표 3〉 대응되는 비교모델 대비 제안모델의 성능 차이(%)

| 제안모델 ID - 비교모델 ID | Accuracy | Precision | Recall | F1-Score | AUC |
|---|---|---|---|---|---|
| OR-T – O-T | 6.87 | 10.94 | -6.05 | 2.72 | 12.12 |
| OR-D – O-D | 14.63 | 16.37 | -1.86 | 8.56 | 20.56 |
| OR-C – O-C | 23.88 | 20.02 | 13.02 | 16.87 | 32.59 |
| OR-TD – O-TD | 15.82 | 8.27 | 18.60 | 13.35 | 17.87 |
| OR-TC – O-TC | 13.13 | 8.65 | 11.16 | 9.80 | 16.12 |
| OR-DC – O-DC | 11.04 | 15.61 | -1.40 | 7.22 | 18.03 |
| OR-TDC – O-TDC | 19.70 | 13.04 | 18.60 | 15.81 | 17.35 |

CIKM 컨퍼런스 발표에서 동일한 데이터셋을 가지고 실험한 결과가 F1-Score 기준 85.84인 것과 비교했을 때, 훨씬 더 우수한 성능임을 간접적으로 확인할 수 있다. F1-Score 뿐 아니라 Accuracy도 0.9045으로 우수하게 산출되었다. 비교모델인 O 모델에서는 가장 성능이 우수한 경우가 O-DC 모델이었는데 Accuracy가 0.7731인 점을 감안하면, 가장 좋은 모델끼리 비교해도 제안모델의 성능이 확실히 더 우수하다고 할 수 있다. 제안모델과 비교모델의 성능 차이를 모델별로 비교한 상세 결과는 다음의 <표 3>에 정리되어 있다.

## 5. 결론

본 연구에서는 유튜브에서 제공하는 관련 동영상 추천이 가짜뉴스의 확산을 촉진시키고 있다는 기존 연구의 결과를 역으로 이용하여, 관련 동영상 정보를 가짜뉴스 탐지의 새로운 정보원으로 활용하는 유튜브 가짜뉴스 탐지 모델을 제안하였다. 제안모델의 성능을 실제 유튜브 가짜뉴스 데이터셋에 적용하여 실증분석한 결과, 원본 동영상의 데이터만으로 가짜뉴스를 판별하는 방법보다, 관련 동영상의 정보를 추가적으로 이용하여 가짜뉴스를 판별하는 방법을 사용했을 때 대부분의 지표가 개선되고 성능이 더 우수함을 알 수 있었다. 이를 통해 관련 동영상의 정보는 가짜뉴스를 판별할 수 있는 중요한 자료로서 가치가 높아 사용할 만한 정보원이라는 것을 확인하였다.

기존에 수행되었던 대다수 유튜브 가짜뉴스 탐지 연구는 콘텐츠 중심의 연구인데 반해, 본 연구는 배경정보를 활용하고자 했다는 점에서 차별화되며, 특히 '관련 동영상' 정보를 배경정보로 활용할 것을 제안한 최초의 연구라는 점에서 학술적으로 의의를 갖는다. 향후 본 연구에서 그 유용성이 확인된 관련 동영상의 정보와 더불어, 기존에 우수했던 콘텐츠 방식의 기법들을 동시에 적용한다면 기존에 콘텐츠만을 사용하였던 연구보다 훨씬 더 좋은 성과를 낼 수 있는 방법을 고안할 수 있을 것이라 판단된다.

이와 같은 학술적 의의를 갖고 있지만 본 연구는 다음과 같은 한계를 갖는다. 첫째, 본 연구는 유튜브에서 얻을 수 있는 다양한 정보 중 제목, 설명, 댓글의 텍스트 정보만을 사용했다는 한계가 있다. 향후 연구에서는 썸네일 이미지나 동영상 자체, 혹은 동영상의 전체 스크립트 등 추가적인 정보를 함께 활용하는 멀티모달(multimodal) 접근법이 연구되어야 할 것으로 판단된다.

둘째, 텍스트를 분석하는 기법은 다양한 방법이 존재하는데 본 연구에서는 그 중 Word2vec의 원리에서 파생된 Doc2vec을 한정하여 사용하였다. Doc2vec 이후에 나온 Transformer 기반의 최신 기술인 BERT나 GPT를 활용한다면 더 좋은 결과를 불러올 수 있으나 기술적, 컴퓨팅 자원의 한계로 인해 실행하지 못했다는 점도 본 연구의 한계이다. 또한 근본이 되는 영상 분석에 대하여는 본 연구자의 지식 및 컴퓨팅 자원의 한계로 인하여 영상 자체를 분석하지 못하였다. 추후 실험에서 이미지 처리를 추가하거나 최신의 언어모델 기술을 사용하고, 나아가 영상까지 분석할 수 있다면 더욱 좋은 결과를 낼 수 있을 것이다.

## 참고문헌(References)

**[국내 문헌]**

김유나. (2021). '한강 의대생 사건' 보도, 언론의 부끄러운 자화상. 관훈저널, 63(3), 79-86.

박성수, 이건창. (2019). 효과적인 가짜 뉴스 탐지를 위한 텍스트 분석과 네트워크 임베딩 방법의 비교 연구. 디지털융복합연구, 17(5), 137-143.

염정윤, 정세훈. (2019). 가짜뉴스 노출과 전파에 영향을 미치는 요인, 한국언론학보, 63(1), 7-45.

이동호, 이정훈, 김유리, 김형준, 박승면, 양유준, 신웅비. (2018). 딥러닝 기법을 이용한 가짜뉴스 탐지. 한국정보처리학회 학술대회논문집, 25(1), 384-387.

이원상. (2019). 허위사실유포에 대한 형법의 대응방안 고찰 - 소위 '가짜뉴스(fake news)'를 중심으로 -. 형사법의 신동향, 62, 35-68.

임혜선. (2023). [초동시각]세상 바꾼 유튜브와 그 다음에 대한 고민. 아시아경제, 2023-02-15 기사 (https://view.asiae.co.kr/article/2023021507252576265)

장윤호, 최병구. (2020). 영상과 텍스트 정보의 결합을 통한 가짜뉴스 탐지 연구 : 유튜브를 중심으로 . 2020 한국경영정보학회 추계학술대회, 231-235.

정정주, 김민정, 박한우. (2019). 유튜브 상의 허위정보 소비 실태 및 확산 메커니즘 생태계 연구: 빅데이터 분석 및 모델링을 중심으로, 사회과학 담론과 정책, 12(2), 105-138.

좌희정, 오동석, 임희석. (2019). 자동화기반의 가짜 뉴스 탐지를 위한 연구 분석, 한국융합학회논문지, 10(7), 15-21.

황용석, 권오성 (2017). 가짜뉴스의 개념화와 규제 수단에 관한 연구 – 인터넷서비스사업자의 자율규제를 중심으로, 언론과 법, 16(1), 53-101.

**[국외 문헌]**

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

Buntain, C., & Golbeck, J. (2017). Automatically identifying fake news in popular twitter threads. In 2017 IEEE international conference on smart cloud (smartCloud) (pp. 208-215). IEEE.

Choi, H., & Ko, Y. (2021). Using Topic Modeling and Adversarial Neural Networks for Fake News Video Detection. In Proceedings of the 30th ACM International Conference on Information & Knowledge Management (pp. 2950-2954).

Choi, H., & Ko, Y. (2022). Effective fake news video detection using domain knowledge and multimodal data fusion on YouTube. Pattern Recognition Letters, 154, 44-52.

Das, M., Singh, P., & Majumdar, A. (2022). Investigating dynamics of polarization of Youtube true and fake news channels. In S. Mukherjee & N. Das (Eds.), Causes and Symptoms of Socio-Cultural Polarization: Role of Information and Communication Technologies (pp. 73-112). Springer Singapore.

Flora, C., & Juliana, G. (2019). YouTube advertises big brands alongside fake cancer cure videos. BBC Trending. Retrieved from https://www.bbc.com/news/blogs-trending-49483681

Le, Q., & Mikolov, T. (2014, June). Distributed representations of sentences and documents. In International Conference on Machine Learning (pp. 1188-1196). PMLR.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

Pan, J. Z., Pavlova, S., Li, C., Li, N., Li, Y., & Liu, J. (2018). Content based fake news detection using knowledge graphs. In The Semantic Web – ISWC 2018: 17th International Semantic Web Conference, Monterey, CA, USA, October 8 – 12, 2018, Proceedings, Part I 17 (pp. 669-683). Springer International Publishing.

Papadopoulou, O., Zampoglou, M., Papadopoulos, S., & Kompatsiaris, Y. (2019). A Corpus of Debunked and Verified User-Generated Videos. Online Information Review, 43(1), 72-88.

Raza, S., & Ding, C. (2022). Fake news detection based on news content and social contexts: a transformer-based approach. International Journal of Data Science and Analytics, 13(4), 335-362.

Sheikhi, S. (2021). An effective fake news detection

method using WOA-xgbTree algorithm and content-based features. Applied Soft Computing, 109, 107559.

Shim, J. S., Lee, Y., & Ahn, H. (2021). A link2vec-based fake news detection model using web search results. Expert Systems with Applications, 184, 115491.

Wynne, H. E., & Wint, Z. Z. (2019, December). Content based fake news detection using n-gram models. In Proceedings of the 21st International Conference on Information Integration and Web-based Applications & Services (pp. 669-673).

Yafooz, W. M. S., Emara, A. M., & Lahby, M. (2021). Detecting fake news on COVID-19 vaccine from YouTube videos using advanced machine learning approaches. In A. Sharma, A. Marques-Pita, & A. S. Ashour (Eds.), Combating Fake News with Computational Intelligence Techniques (pp. 421-447). Springer.

Abstract

# Fake News Detection on YouTube Using Related Video Information

Junho Kim* · Yongjun Shin** · Hyunchul Ahn***

As advances in information and communication technology have made it easier for anyone to produce and disseminate information, a new problem has emerged: fake news, which is false information intentionally shared to mislead people. Initially spread mainly through text, fake news has gradually evolved and is now distributed in multimedia formats. Since its founding in 2005, YouTube has become the world's leading video platform and is used by most people worldwide. However, it has also become a primary source of fake news, causing social problems. Various researchers have been working on detecting fake news on YouTube. There are content-based and background information-based approaches to fake news detection. Still, content-based approaches are dominant when looking at conventional fake news research and YouTube fake news detection research. This study proposes a fake news detection method based on background information rather than content-based fake news detection. In detail, we suggest detecting fake news by utilizing related video information from YouTube. Specifically, the method detects fake news through CNN, a deep learning network, from the vectorized information obtained from related videos and the original video using Doc2vec, an embedding technique. The empirical analysis shows that the proposed method has better prediction performance than the existing content-based approach to detecting fake news on YouTube. The proposed method in this study contributes to making our society safer and more reliable by preventing the spread of fake news on YouTube, which is highly contagious.

**Key Words** : Fake News, YouTube, Doc2vec, Convolutional Neural Network, Related Video

Received : May 15, 2023 Revised : May 15, 2023 Accepted : May 30, 2023
Corresponding Author : Hyunchul Ahn

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

# 저 자 소 개

**김준호**
현재 국민대학교 비즈니스IT전문대학원 석사과정으로 재학 중이다. 국민대학교에서 사법학전공 법학사를 취득하였다. 주요 관심 분야는 비즈니스 애널리틱스(BA), 추천 시스템, 텍스트 마이닝, 딥 러닝, 머신 러닝을 활용한 각종 이상 탐지 등이다.

**신용준**
현재 이노베이션 아카데미에서 운영하는 42SEOUL의 학생으로서 컴퓨터과학에 대해 더 깊게 공부하고 있다. 강원대학교에서 컴퓨터공학 공학사를 취득하였다. 주요 관심 분야는 그래픽 처리, 데이터 수집이다.

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