요소 정보 활용을 통한 가짜 뉴스 탐지

한상도[O], 이근배[†]

포항공과대학교 컴퓨터공학과[O,†], 포항공과대학교 인공지능대학원[†]
{hansd, gblee}@postech.ac.kr

Fake news detection via news elements

Sangdo Han[O], Gary Geunbae Lee[†]
Dept. of Computer Science and Engineering, Pohang University of Science and Technology[O,†]
Graduate School of Artificial Intelligence, Pohang University of Science and Technology[†]

요 약

본 연구에서는 가짜 뉴스 탐지를 위한 데이터를 구축하고, 내용 기반의 탐지를 위한 시스템을 제안하였으며, 뉴스의 각 요소 정보가 탐지 성능에 미치는 영향을 확인하였다. 이는 기존의 내용 기반 가짜 뉴스 탐지 방법론들의 단점을 보완할 뿐 아니라 뉴스의 요소 정보가 진위 판별에 미치는 영향을 확인하기 위함이었다. 이를 위해 직접 구축한 뉴스 데이터의 제목과 본문을 따로 인코딩하여 판별하였고, 각 요소를 배제한 실험을 통해 뉴스 제목이 가장 중요한 요소 정보임을 확인하였다. 결과적으로 자극적인 제목으로 이목을 끌려는 가짜 뉴스의 속성을 정량적으로 확인할 수 있었다.

주제어: 가짜 뉴스 탐지, 딥 러닝, 뉴스 요소 정보

1. 서론

가짜 뉴스는 혐오를 조장하고, 사회적 비용을 초래하기 때문에 다양한 가짜 뉴스 탐지 기법들이 연구 되어왔다. 특히 뉴스 내용 기반의 탐지 기법들은 가짜 뉴스에 즉각적인 대응이 가능하기 때문에 여러 시도가 이루어져 왔다. 그러나 기존에 구축된 데이터와 방법론들은 뉴스의 진위를 판별하려는 목적과 부합하지 않는 면이 있었다. 본 연구는 가짜 뉴스 탐지를 위해 데이터를 직접 구축하였으며, 내용 기반 판별을 위한 딥 러닝 기반의 시스템을 설계하였다. 더 나아가 뉴스의 내용 정보를 제목과 본문의 요소 정보로 분리하고, 각 요소 정보를 따로 인코딩하는 방법론을 제안하였다. 실험 결과, 제목이 자극적이어야 하는 가짜 뉴스의 특성 상, 제목을 활용할 경우에 F1 점수가 가장 높았으며, 본문을 함께 활용할 경우 정확도 향상에 도움이 되는 것을 확인할 수 있었다.

2. 관련 연구

가짜 뉴스 탐지는 최근 활발하게 연구되고 있는 분야이다. 본 연구의 방법론 소개에 앞서 데이터 구축 및 방법론에 대한 관련 연구를 소개하고자 한다.

대표적인 가짜 뉴스 데이터로는 LIAR[5]와 FakeNewsNet(FNN)[4]데이터가 있다. 이 두 데이터는 모두 PolitiFact.com이라는 가짜 뉴스 판별 서비스를 제공하는 웹 페이지를 추출하여 구축했다는 공통점이 있다. LIAR 데이터는 딥 러닝 방법론을 적용할 만큼의 크기(12836 데이터)를 제공하지만, 판별의 대상이 되는 뉴스 기사(article) 대신 주장(claim)을 제공한다는 한계가 있었으며, FNN의 경우는 뉴스 기사를 제공하지만 그 규모가 작아(1056 데이터) 딥 러닝 방법론을 적용하기에는 한계가 있었다. 본 연구에서는 기존의 문제점들을 해소하기 위해 뉴스 기사를 충분한 크기로 새로 구축하였다.

가짜 뉴스 방법론은 크게 사회 관계망 서비스를 활용한 탐지 방법[2]과 뉴스 내용 기반 탐지 방법[6]의 두 가지로 이루어져 왔다. 사회 관계망 서비스를 활용하는 경우, 뉴스에 대한 사람들의 반응 정보를 활용하여 진위 여부를 판단하는 방법으로, 판별 정확도가 상당히 높은 방법이다. 그러나 이 방법은 기사가 게재된 후 사용자들의 반응을 기다려야 하기 때문에 판별 속도가 늦어 대응이 느리다는 단점이 있다. 뉴스 내용 기반의 탐지 방법은 기존의 연구들은 데이터 양이 적어 딥 러닝 방법을 시도해보지 않았거나[6], 딥 러닝 방법을 시도한 방법들은 뉴스 기사가 아닌, 기사에서 발췌한 주장을 근거로 판단했거나, 혹은 적은 양의 데이터에 딥 러닝 방법론을 적용하여 충분히 훈련하지 못했다는 한계가 있었다[1]. 본 연구에서는 이러한 단점들을 보완하기 위해 딥 러닝을 적용할 만큼의 뉴스 기사를 수집하여 기사 자체의 진위 여부를 판단하도록 진행하였다. 추가적으로, 뉴스 기사를 기반으로 진위 여부를 판단하는데 있어 요소 정보의 중요도를 확인할 수 있도록 실험을 진행하였다.

3. 데이터 구축

본 연구에서는 해외에서 제공하는 가짜 뉴스 판별 서비스인 PolitiFact.com 사이트를 직접 추출하여 구축하

† corresponding author

였다. PolitiFact.com 사이트에서 제공하는 서비스 정보는 주장(claim), 판단(label), 논설(article), 근거 자료(source)의 구조로 이루어져 있다(그림 1). 우리는 뉴스 기사의 진위 여부를 판단하고자 하는 본래 목적에 부합할 수 있도록, 근거 자료를 데이터로서 수집하였다. 본 뉴스 구축 방법을 통해 기존의 PolitiFact.com을 추출한 다른 FNN이나 LIAR데이터가 뉴스가 아닌 주장을 추출하였다는 점을 보완했다. 추가로, 주장의 내용에 가장 부합하는 뉴스만을 추출하기 위해 주장 문장과 뉴스 기사 간의 유사도가 가장 높은 뉴스를 추출하였다. 이를 위하여 주장 문장과 뉴스 기사는 TF-IDF 벡터로 산출한 후 벡터 간 유사도가 가장 높은 뉴스 기사를 선택하는 방식을 택하였다.

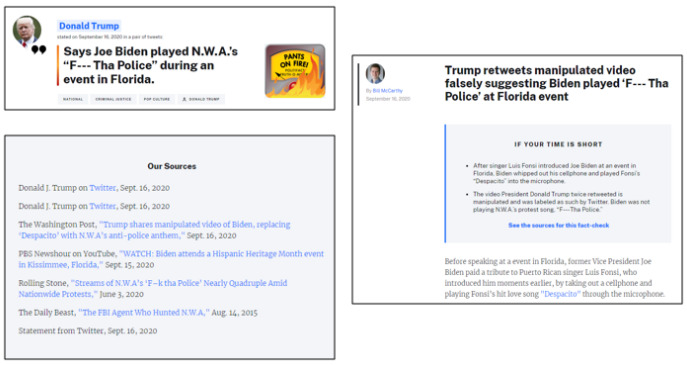

그림 1. PolitiFact.com에서 제공하는 주장과 판단 정보(왼쪽 위), 논설 정보(오른쪽), 근거자료(왼쪽 아래) 정보

본 연구는 PolitiFact.com 에서 추출한 데이터를 이진 분류 문제로 간주하였는데, 실제 데이터는 label 정보가 6가지 분류로 이루어져 있다(pants-fire, false, mostly-false, half-true, mostly-true, true). 우리는 3가지 분류(pants-fire, false, mostly-false)를 거짓으로, 3가지 분류를 사실(half-true, mostly-true, true)로 분류하여 이진 분류 데이터로 간주하였다. 데이터 예시는 표 1로 확인할 수 있다.

표 1 데이터 예제

속성	데이터
고유번호	10413
진위	true
제목	Did President Obama save the auto industry?
본문	Even the most casual viewer of the Democratic convention would get the point: President Barack Obama saved the American auto industry. ...

본 연구에서 추출한 데이터는 PolitiFact.com 사이트에 게재된 2009년 10월부터 2020년 5월까지의 진위 판별 문서를 대상으로 하였다. 구축된 데이터의 크기는 총 12547 뉴스로, 훈련/검증/테스트 셋으로 나누었다(표 2).

표 2 구축된 뉴스의 통계정보

	train	dev	test
뉴스 수	10130	1247	1170

4. 방법 및 결과

우리는 가짜 뉴스 분류를 위해, 뉴스 기사를 인코딩할 모델을 설계하였다. 또한 뉴스의 각 요소 정보(제목, 본문)이 진위 판별에 미치는 영향을 확인해 보기 위해 제목과 본문을 분리하여 비교하였다. 뉴스의 제목을 인코딩 하기 위해 BERT를 활용하였으며, 본문의 경우 제목을 인코딩하는 BERT와는 독립적으로 CNN을 활용하였다. 이러한 선택을 한 이유는 두 가지가 있는데, 첫째, 본문은 제목보다 상대적으로 길이가 길고 데이터 간 길이 편차가 크기 때문에 BERT를 사용할 경우 연산량이 기하급수적으로 늘어날 뿐 아니라 데이터 마다 상당히 다른 패딩 길이가 인코딩 성능에 영향을 줄 것을 염려하였다. 둘째, 뉴스의 특성 상 본문에 다양한 맥락의 정보들이 담겨 있기 때문에 하나의 맥락 정보로 표현하기 보다는 기사에 담겨있는 다양한 단위 정보가 소실되지 않게 표현할 수 있도록 인코딩하기 위함이었다. 보통 BERT의 훈련의 대상이 되는 위키피디아 문서의 경우엔 문서 전체가 정보의 통일성을 담고 있으나, 뉴스는 찬성과 반대 의견이 함께 담기기도 하기 때문이다. 제목 인코딩 벡터와 본문 인코딩 벡터는 결합되어 뉴스 임베딩 벡터로 최종 산출되었으며, 이는 순방향 신경망(FC)을 적용하여 진짜/가짜 뉴스를 판별하도록 설계하였다. 본 연구에서 제안하는 가짜 뉴스 탐지의 전체 흐름도는 그림 2와 같다.

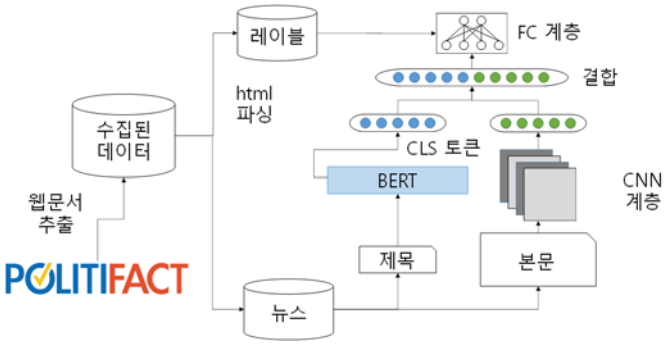

그림 2. 제목-본문 분절 가짜 뉴스 탐지 흐름도

본 모델의 BERT는 사전 훈련된 BERT 모델을 사용하였으며, CNN과 FC계층은 새로 훈련하였다. CNN 계층은 3,4,5 사이즈의 커널을 100개씩 활용하였으며 dropout ratio는 0.1을 사용하였다. CNN에 입력하는 단어 임베딩은 사전 훈련된 GloVe 임베딩[3]을 사용하였으며, fine-tuning 가능하도록 훈련하였다. FC계층은 하나의 hidden layer를 가지는 간단한 구조를 사용하였다.

가짜 뉴스 탐지 실험 결과는 표 3과 같다. 우리가 구축한 데이터를 대상으로 한 기준 성능이 없기 때문에,

뉴스 기사의 진위 여부를 무작위로 판단하는 경우를 기준 성능으로 두었다. 실험 결과는 가짜 뉴스를 positive label로 간주한 f1 점수는 제목만을 활용할 때 가장 높게 나왔으며, 정확도(accuracy)는 제목과 본문을 함께 활용할 때 가장 높았다.

표 3 가짜 뉴스 탐지 결과

방법론	F1	precision	recall	accuracy
무작위	0.5050	0.5128	0.4975	0.4974
제목+본문	0.5995	**0.6542**	0.5534	**0.6575**
제목	**0.6334**	0.6157	0.6522	0.6502
본문	0.5274	0.5113	**0.6683**	0.5040

5. 결론

본 연구에서는 가짜 뉴스를 자동으로 탐지하는 시스템을 위하여 직접 데이터를 구축하고 딥 러닝 방법으로 탐지를 해 보았다. 특히, 뉴스가 가지는 두 가지 요소인 제목과 본문을 나누어 가짜 뉴스 탐지에 활용하고 그 결과를 분석해 보았다. 실험 결과, 가짜 뉴스를 탐지하기 위해 가장 중요한 정보는 제목이며, 이는 사람들의 시선을 끌기 위해 제목을 최대한 자극적으로 선정하고자 하는 가짜 뉴스들의 특성과도 부합되었다. 실험 결과에 따르면 본문 만을 활용하는 경우에는 효과적인 탐지가 불가능 했지만, 보조 역할로 활용할 경우 정확도 향상에는 도움이 됨을 확인할 수 있었다. 차후 연구로는 가짜 뉴스 탐지에 중요하다고 언급되는 출처 정보를 활용한 연구로 확장할 계획이다.

Acknowledge

이 논문은 2020년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(No. 2019-0-01906, 인공지능대학원지원(포항공과대학교))을 받아 수행된 연구임

참고문헌

[1] Jwa, Heejung et al, exBAKE: Automatic Fake News Detection Model Based on Bidirectional Encoder Representations from Transformers (BERT), Applied Sciences, vol 9, no 19, pp. 4062-4071, 2019.
[2] Ma, Jing et al, Rumor detection on twitter with tree-structured recursive neural networks, Association for Computational Linguistics, 2018.
[3] Pennington, Jeffrey et al, GloVe: Global vectors for word representation, Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pp.1532-1543, 2014.
[4] Shu, Kai et al, 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, pp. 171-188, 2020.
[5] Wang, William Yang, “Liar, Liar Pants on Fire”: A New Benchmark Dataset for Fake News Detection, Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pp. 422-426, 2017.
[6] Zhou, Xinyi et al, Fake news early detection: A theory-driven model, Digital Threats: Research and Practice, vol 1, no 2, pp.1-25, 2020.