

딥러닝 기반 가짜뉴스 탐지 시스템에서 뉴스 카테고리의 영향 분석

한윤진*, 김성한**, 김근형*

*동의대학교, **한국전자통신연구원

qkr030@naver.com, sh-kim@etri.re.kr, geunkim@deu.ac.kr

Analysis of the effect of News Categories in Deep Learning-based Fake News Detection Systems

YoonJin Han*, Sunghwan Kim**, Geun-Hyung Kim*

*Dong-eui Univ., **ETRI

요약

소셜미디어(SNS)가 보편화되어 개인 방송 채널이 증가하면서 거짓 정보를 언론 보도 형식으로 위장한 가짜뉴스가 꾸준히 증가하고 있어 사회적 비용을 많이 발생시킬 것이다. 최근 딥러닝 기술을 활용하여 가짜뉴스를 자동으로 탐지하는 연구가 활발히 진행되고 있다. 본 논문에서는 가짜뉴스 탐지 모델 개발을 위해 설계한 메타 데이터 체계에 따른 가짜뉴스의 분류가 가짜뉴스 탐지 모델의 성능에 미치는 영향을 분석하였다.

I. 서론

한국언론진흥재단은 정치·경제적 이익을 얻으려 의도적으로 언론 보도 형식으로 유포된 거짓 정보를 가짜뉴스로 정의하고 가짜뉴스를 정보 형태에 따라 동영상, 오디오, 이미지, 텍스트로 구분하며 정보 성격에 따라 허위정보, 거짓 정보, 오인정보, 패러디, 루머로 분류한다[1]. 가짜뉴스 탐지 방법은 크게 비기술적 방법, 기술적 방법, 하이브리드 방법으로 분류한다[2].

본 논문에서는 기술적 가짜뉴스 탐지 방법인 딥러닝 기반 가짜뉴스 탐지 모델의 성능과 뉴스 데이터의 카테고리 분류의 관계 (뉴스 카테고리가 모델 성능에 미치는 영향)를 LSTM(Long Short-Term Model) [3] 기반의 딥러닝 네트워크를 사용하여 분석하였다.

II. 본론

본 논문에서는 다양한 미디어의 가짜뉴스를 콘텐츠 또는 맥락의 기반으로 검출하는 인공지능 모델을 개발할 때 활용되는 뉴스 데이터를 분류하기 위해 정립한 메타 데이터 체계[4]를 바탕으로 Kaggle[5]의 뉴스 데이터를 분류하여 뉴스 데이터의 분류가 딥러닝 기반 가짜뉴스 탐지 모델의 성능에 미치는 영향을 분석하였다. 정립한 메타 데이터 체계는 미디어에 공통이고 필수적인 메타 데이터 요소와 미디어 특화되고 선택적인 메타 데이터 요소로 구분하였고 공통적인 메타 데이터 요소는 카테고리, 제목, 본문 또는 콘텐츠에 해당하는 내용 또는 링크, 뉴스 게재 일자, 뉴스 게시자, 출처가 있다.

본 연구에서는 가짜뉴스 탐지 모델을 개발 및 활용할 때 뉴스 데이터를 메타 데이터 중 공통요소의 중요성을 강조하기 위해 실험을 진행하였다. 우리는 실험을 통해 메타 데이터의 공통요소 중 카테고리 요소가 가짜뉴스 탐지에 얼마나 영향을 미치는지 분석하였다. 실험에 사용한 인공지능 모델은 심층 신경망(DNN: Deep Neural Network)중 하나인 LSTM 네트워크로 구성하였고, 학습 및 테스트

트용 뉴스 데이터는 Kaggle의 뉴스 데이터 셋에서 진짜뉴스 21,417개와 가짜뉴스 23,481개를 추출하여 사용되었다. 검증용 뉴스 데이터는 학습용 뉴스 데이터의 일부를 활용한다. 사용된 Kaggle의 뉴스 데이터는 제목, 본문, 카테고리, 그리고 게시일자 정보로 이루어져 있다.

첫 번째 실험에서는 카테고리 상관없이 진짜뉴스 데이터 5,000개와 가짜뉴스 데이터 5,000개를 무작위로 추출하였다. 추출한 10,000개의 뉴스 데이터는 'news'로 분류된 데이터가 4,516개, 'worldnews'로 분류된 데이터가 2,410개, 'politics'로 분류된 데이터가 1,469개, 'left-news'로 분류된 데이터가 942개, 'government News'로 분류된 데이터가 330개, 'middle-east'로 분류된 데이터가 171개, 'us_news'로 분류된 데이터가 162개로 구성되었다. 첫 번째 실험에서는 학습 데이터 8,000개와 테스트 데이터 2,000개로 나누어 사용하였으며, 학습 과정에서 학습 데이터의 20%를 검증 데이터로 사용하였다.

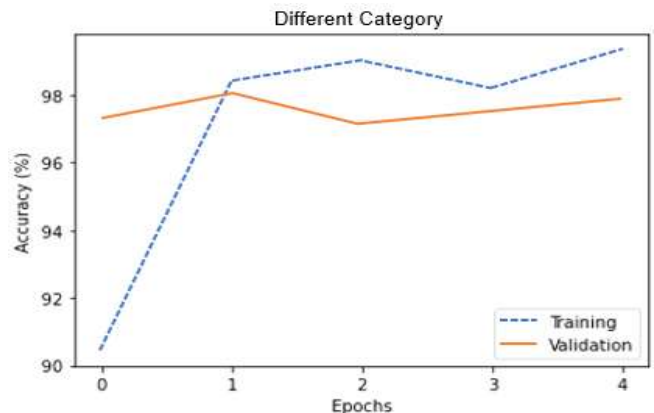


그림 1. 카테고리 분류가 없는 뉴스 활용 학습 결과.

두 번째 실험에서는 동일한 카테고리에 속한 진짜뉴스 데이터 5,000개와 가짜뉴스 데이터 5,000개를 무작위로 추출하였다. 10,000개의 뉴스 데이터는 'news'로 분류된 데이터로만 구성하였다. 첫 번째 실험과 같은 비율로 데이터를 분리하여 실험에 사용하였으며 두 가지 실험 모두 5번의 epoch를 반복하여 학습하였다.

그림 1은 카테고리를 분류하지 않은 실험의 결과를 나타낸다. 카테고리 종류에 상관없이 데이터를 추출하여 사용한 첫 번째 실험 결과(그림 1)를 보면 테스트 데이터는 epoch를 반복할수록 학습 정확도는 증가하나 검증 데이터는 정확도가 0.98을 넘지 않고 유지된다.

동일 카테고리에 속한 뉴스 데이터를 기반으로 실험한 결과는 그림 2와 같다. 여러 카테고리에 속한 뉴스 데이터를 기반으로 실험한 결과보다 동일 카테고리에 속한 뉴스 데이터를 기반으로 한 실험이 가짜뉴스 탐색 모델의 정확도가 높다.

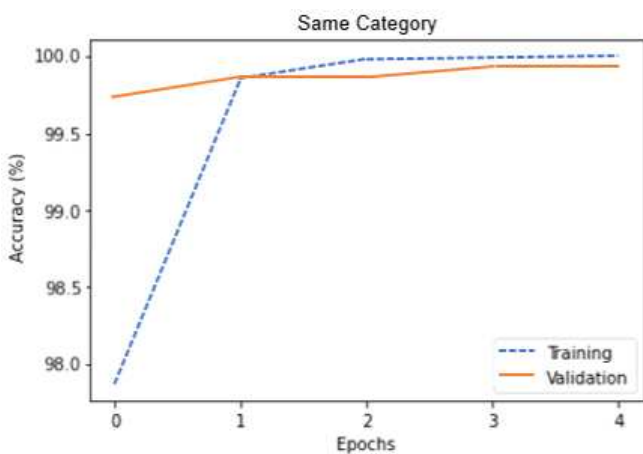


그림 2. 동일 카테고리로 분류된 뉴스 활용 학습 결과.

카테고리가 다른 뉴스 데이터를 이용한 실험 결과(그림 1)와 달리 동일한 카테고리로 분류된 뉴스 데이터를 이용한 두 번째 실험 결과(그림 2)를 보면 테스트 데이터는 epoch를 반복할수록 학습 정확도와 검증 정확도가 모두 증가한다.

학습된 두 모델에 테스트 데이터를 사용한 테스트를 하면 첫 번째 학습 모델의 가짜뉴스 탐지 정확도는 47.36%, 두 번째 학습 모델의 가짜뉴스 탐지 정확도는 98.68%이다. 테스트에 사용된 2,000개의 뉴스 데이터는 'news' 카테고리 데이터가 594개, 'worldnews' 카테고리 데이터가 469개, 'politics' 카테고리 데이터가 367개, 'left-news' 카테고리 데이터가 359개, 'government News' 카테고리 데이터가 150개, 'middle-east' 카테고리 데이터가 31개, 'us_news' 카테고리 데이터 30개로 이루어졌다. 그러나 학습에 사용된 10,000개의 뉴스 데이터 중 4,516개가 'news' 카테고리에 해당되어 테스트 시 가짜뉴스 탐지의 정확도가 낮았다.

실험 결과 학습 데이터를 세부 카테고리로 분류하지 않고 학습시킨 가짜뉴스 탐지 모델은 다양한 뉴스 데이터에 대해 가짜뉴스를 탐지할 때 정확도가 낮았다. 실험 결과로부터 유사한 특징(예: 카테고리)을 가지는 뉴스 데이터로 학습시킨 가짜뉴스 탐지 모델에 동일한 카테고리에 속한 뉴스에 대해서만 뉴스의 진위를 판단할 때 정확성이 높다고 추론할 수 있다.

인공지능 모델은 설계 시 설정한 조건 범위 및 지식의 범위 내에서만 잘 동작하는데 모델이 작동하도록 설계 또는 고려되지 않은 경우의 데이터에 대해서는 모델의 정확도가 낮아지는데 이와 같은 인공지능 모델의 지식 한계[6]를 식별하고 선언하는 것이 필요하다.

III. 결론

본 논문에서는 가짜뉴스 탐지 모델 개발을 위해 정립한 메타 데이터 체계에 따른 가짜뉴스의 카테고리 분류가 가짜뉴스 탐지 모델의 성능에 미치는 영향을 분석하였다. 가짜뉴스가 다양한 미디어 기반으로 생성되고 있어 가짜뉴스를 분류할 수 있는 메타 데이터 체계가 요구되며 메타 데이터의 카테고리화 인공지능 모델의 지식 한계가 가짜뉴스 탐지 모델의 정확도에 미치는 영향을 분석하기 위해서 뉴스 데이터를 카테고리 분류 없이 학습 데이터로 활용한 가짜뉴스 탐지 모델과 하나의 카테고리에 속한 뉴스 데이터를 학습 데이터로 활용한 가짜뉴스 탐지 모델의 성능을 분석하였고 카테고리별 가짜뉴스 탐지 모델을 개발하는 것이 적합하다는 결론을 도출하였다.

가짜뉴스 탐지 모델을 활용하기 위해 뉴스를 메타 데이터 체계에 따른 올바른 분류가 가짜뉴스 탐지 모델의 성능에 영향을 미치기 때문에 뉴스 데이터의 자동 분류 기능이 요구되며 인공지능 모델의 지식 한계를 규정하고 지식 한계를 벗어나는 것에 대해서 결정을 유보할 수 있어야 한다.

감사의 글

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 한국 연구재단의 지원을 받아 수행된 연구임(No. NRF-2021R1F1A1047573).

참고 문헌

- [1] 좌희정, 오동석, 임희석, "자동화기반 가짜 뉴스 탐지를 위한 연구 분석", 한국융합학회논문지, 제10권, 7호, pp15-21, 2019.
- [2] 윤영석, 엄태원, 이재영, 이현우, 허재두, "페이스 뉴스 탐지 기술 동향과 시사점", 주간기술동향, 정보통신기술진흥센터, Oct. 2017.
- [3] S. Hochreiter, J. Schmidhuber, "Long Short-Term Memory," Neural Computation, 9(8), pp. 1735-1780, 1997.
- [4] 김근형, 김성환, 한윤진, "인공지능 기반 가짜뉴스 탐지 시스템을 위한 뉴스 메타 데이터 연구 분석," 한국멀티미디어학회 추계학술발표대회논문집, 제 23권 2호, 2020.
- [5] Kaggle 홈페이지, <https://www.kaggle.com>.
- [6] P. J. Phillips, C. A. Hahn, P. C. Fontana, D. A. Broniatowski, M. A. Przybocki, "Four Principles of Explainable Artificial Intelligence," NIST Report, Draft NISTIR 8312, August 2020.