

# 딥러닝을 활용한 코로나 19 이후 종교 인식 변화 분석

구민구<sup>o</sup>

김용환<sup>1</sup>

노기섭<sup>2</sup>

청주대학교 {인공지능소프트웨어전공<sup>o2</sup>, 문헌정보학전공<sup>1</sup>}

## Analysis of Perception Changes in Religion after COVID-19 using Deep Learning

Min-ku Koo<sup>o</sup>

Yonghwan Kim<sup>1</sup>

Giseop Noh<sup>2</sup>

Department of {Artificial Intelligence Software<sup>o2</sup>, Library and Information Science<sup>1</sup>}

Cheongju University

koomk97<sup>o</sup>, kimyonghwan<sup>1</sup>, kafa462 {@cju.ac.kr}

### 요 약

코로나 19 는 다양한 형태로 영향을 미치고 있으며 삶의 형태를 변화시키고 있다. 본 논문에서는 국내 유입으로 인해 변화된 종교에 대한 국민의 인식을 데이터 기반으로 분석하였다. 종교와 관련된 키워드를 ‘기독교’, ‘천주교’, ‘불교’, ‘신천지’, ‘종교’로 선정하고 관련 뉴스 기사의 댓글을 수집하였다. 댓글의 감성값을 산출하기 위하여 일부 댓글의 감성값에 대한 레이블링을 진행하고, 딥러닝 모델을 통해 값을 산출하였다. 본 연구에서는 데이터 수집을 위한 프로그램을 직접 구현하여 종교에 따른 기사 및 댓글을 수집하고, 각 종교별 감성을 딥러닝을 이용하여 분석하였다. 추가적으로 월별 댓글 수를 분석하였다. 본 연구를 통해 종교에 대한 전체 키워드에서 코로나 19 이후 부정적 감성이 증가함을 발견하였으며 종교별 감성변화에 연관성에 대한 결과를 도출하였다.

## 1. 서 론

2020 년 1 월, 전 세계로 퍼진 코로나 19 (COVID-19)로 인해 사람들의 일상생활은 많은 영향을 받았다. 또한, 2020 년 2 월 18 일 특정 종교집단으로부터 집단감염(본 논문에서는 ‘신천지’ 관련 사항을 분석함)이 시작되었으며, 2020 년 6 월 또다른 집단 감염, 그리고 2020 년 8 월 광복절 관련 대규모 집회 등이 발생하였다. 이와 같은 사회적 이벤트는 코로나 19 의 확산을 촉진시켰으며 확진자 수가 급격히 증가함을 확인할 수 있다. 종교와 관련된 사회 이슈로부터 언론은 종교에 대한 다양한 온라인 정보를 생성하였다. 기계학습(Machine Learning, 본 논문에서는 Fully Connected Layer 기반 딥러닝 학습 적용) 기반 감성 분석(Sentiment Analysis)을 진행 통해 코로나 19 가 대한민국의 종교계에 대한 사람들의 인식에 어떤 영향을 미쳤는지 분석하였다. 본 논문에서는 온라인 데이터 수집 및 처리, 딥러닝을 이용한 감성 분석을 이용하여 코로나 19 로부터 파급된 종교의 영향력에 대하여 분석을 시도하였다. 종교 관련 언급량, 감성, 댓글 빈도수 등을 분석한 결과 코로나 19 이후 종교에 대한 한국인들의 인식과 관심 주제의 변화를 데이터 기반으로 수치화하여 확인할 수 있었다(그림 1 참조).

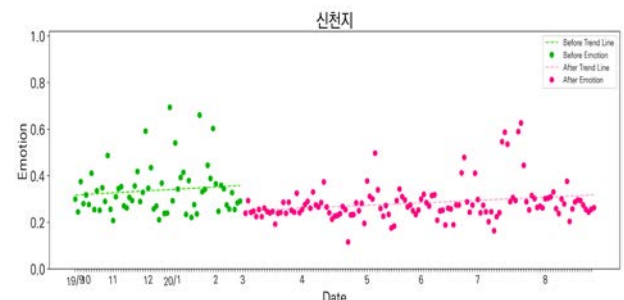


그림 1. ' 20. 2 월 전후 신천지 관련 감성변화

## 2. 연구 방법

종교에 대한 한국인들의 인식을 파악하기 위해 포털사이트 점유율이 높은 네이버와 다음을 활용하여 데이터 수집을 진행하였다. 댓글 수집을 위한 뉴스의 검색 키워드는 ‘기독교’, ‘불교’, ‘천주교’, ‘신천지’, ‘종교’까지 총 5 가지로 선정하였다. 그리고 코로나 19 유입 날짜를 기준으로 이전 5 개월, 이후 5 개월 기간에서 15,291 개의 기사와 748,020 개의 댓글 데이터를 수집하였다.

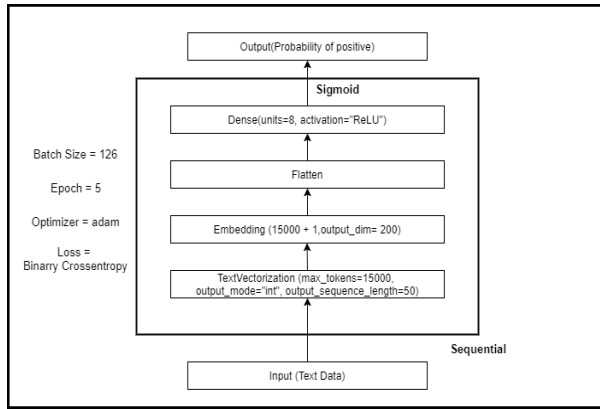


그림 2. 종교 관련 댓글의 딥러닝 감성 학습 구조

감성분석을 위해서 수집된 데이터 중에서 974 개의 데이터에 대해 Human Inspection 방식을 통해 감성값을 Labeling(긍정일 경우 1, 부정일 경우 -1로 판정) 하였다. Human Inspection을 통한 Labeling Data는 전체 데이터의 0.13%로 감성 학습을 위한 데이터로는 부족하다. 이를 보완하기 위해 네이버 영화 리뷰 데이터[1]를 확보하여 확보한 데이터와 합쳐서 추가 학습을 진행하였다. [1]에서 제공하는 데이터는 5만 개의 학습데이터와 15만개의 테스트 데이터로 구성되어 있다.

수집된 데이터는 이후 KoNLPy의 Okt 라이브러리를 활용하여 데이터 전처리(Data Preprocessing)를 하였다. 감성분석 모델을 구축하기 위해 Keras 프레임워크를 이용하여 딥러닝 네트워크를 생성하였다. TextVectorization과 Embedding 내장 함수를 활용하여 문자열 데이터를 벡터화시켰으며, ReLU 활성화 함수를 이용한 은닉층을 추가하고, 출력층에는 sigmoid 함수를 이용하였다. 손실 함수(Loss function)는 Binary Crossentropy, 최적화 함수(Optimizer)는 Adam으로 설정한 다음 모델을 훈련시킨 결과, 훈련 정확도는 약 95%, 손실 값은 약 0.118이며, 테스트 정확도는 82%, 손실 값은 0.61이다.

추론(Inference)과정은 모델에 댓글을 입력하면 0과 1 사이의 값으로 감성값이 나타나며, 0에 가까울수록 부정적인 감성, 1에 가까울수록 긍정적인 감성을 의미한다. 모든 댓글을 생성된 모델에 적용하여 감성을 Labeling 하였다. 키워드별 기간에 따른 감성의 차이가 유의미한지 확인하기 위해 95% 신뢰도를 기준으로 T-test를 수행하였다. 검정결과 모든 키워드에서 코로나 발생 이전 기간과 이후에 유의미한 차이가 있음을( $p < .05$ ) 확인하였다.

날짜당 감성값은 기사 하나의 댓글 개수의 비율을 고려하여 가중 평균을 이용하였다. 이후 전체 기간에서 해당 날짜의 감성값을 반영하기 위해 IMDB 투표수를 고려한 영화 순위 예측 방식을 적용한다[2]. 또한, 키워드별 기간에 따른 단어 빈도 분석을 위하여 명사만을 추출하여 빈도순 기준으로 100개 단어를 분석하였다.

### 3. Analysis & Findings

코로나 발생 이전보다 이후에 종교에 대한 감성은 부정적으로 변화하였다(신천지 -17.8%, 기독교 -10.8% 감성값 감소). 모든 키워드에서 코로나 발생 이후 기간의 댓글 개수가 많았다. 코로나 발생 이후 기간에서는 ‘교회’, ‘예배’ 등 종교 시설과 활동에 대한 단어 순위가 상승했다.

키워드 ‘기독교’, ‘신천지’에서는 집단감염 이슈가 발생한 시점에서 부정 감성이 증가하였다. 특이한 점은 ‘천주교’, ‘불교’ 관련 키워드는 언급량이 적고 연관 사회 이슈가 없음에도 불구하고 ‘기독교’ 및 ‘신천지’ 관련 부정 감성 감소의 경향을 따르는 것을 확인하였다(천주교 -24.3%, 불교 -23.6% 감소). ‘천주교’는 기독교보다 긍정적이었으나 부정 감성으로 크게 전이되었다. 이는 타 종교간 감성 전이가 발생한 것으로 추측된다. 이에 대한 데이터 기반 연구는 향후 연구로 남긴다. ‘신천지’ 관련 감성 변화를 그림 1에 제시하였다(잔여 키워드 감성 변화는 지면 제약으로 생략한다.) 코로나 19는 온라인 공간에서 종교에 대한 관심도는 높였지만, 이전에도 부정적이었던 종교에 대한 감성을 모든 종교관련 분야로 좀더 부정적으로 전이시킨 것으로 파악되었다.

### 4. 결론

본 논문에서는 코로나 19 이후 기간에서 종교의 인식 변화를 키워드 중심 온라인 뉴스 데이터를 기반으로 딥러닝을 활용해 분석하였다. 종교와 관련된 총 5가지 키워드 모두 코로나 발생 이후 기간에서 이전 기간보다 부정 감성이 증가하였다. 본 연구에서는 댓글 감성 레이블링을 Human Inspection으로 진행하다 보니 주관적인 측정이 될 수 있다는 점의 한계점으로 볼 수 있으나 현재까지 가장 정확도가 높은 방법이므로 수용 가능하다. 본 연구의 한계점으로는 한국인의 44%가 종교인이라는 점을 고려 시 일부 종교의 인식 변화가 사회현상 분석의 대표성을 완전히 담보하지 못한다는 점이다. 향후 연구는 종교 이외 분야를 포함하여 데이터기반 분석을 추진할 예정이다.

### Acknowledgement

이 성과는 2019년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. NRF-2018R1C1B5083091).

### 5. 참고 문헌

- [1] Naver sentiment movie corpus v1.0(2016), 12.Jan.2021, from <https://github.com/e9t/nsmc>, accessed Jan 30, 2021.
- [2] Manish Sahay, How does IMDb's rating system work, accessed Jan. 19, 2021, from <https://www.quora.com/How-does-IMDBs-rating-system-work>