

<http://dx.doi.org/10.17703/JCCT.2022.8.1.453>

JCCT 2022-1-52

딥러닝을 활용한 출산율 감소에 따른 모병제 인식 변화 분석

An Analysis of Volunteer Military System Perception Changes with Decreasing Fertility Rates using Deep Learning

구민구*, 박지용**, 이현무***, 노기섭****

Minku Koo*, Jiyong Park**, Hyunmoo Lee***, Giseop Noh****

요약 한 나라의 출산율 감소는 생산가능인구가 감소하고, 인구구조 고령화에 따른 저축률 저하로 자본축적이 줄어들어 경제성장이 둔화 등의 문제가 발생한다. 현재 대한민국에서는 만 18세 이상의 모든 남성이 병역의 의무를 부과하고 있는 징병 제도를 시행하고 있다. 하지만 출산율 감소로 인해 모병 제도로의 전환이 사회적 이슈로 불거지고 있다. 본 논문에서는 출산율이 1 미만으로 떨어진 2018년부터 모병제에 대한 사회 인식을 분석하고자 '모병제' 키워드를 통해 검색된 뉴스 기사와 댓글을 수집하였다. 수집된 댓글 중 일부에 대해 레이블링을 진행하였고, 딥러닝 모델을 통해 감성 수준을 산출하였다. 본 연구를 통해 출산율 저하에 따라 모병제 전환에 대한 인식이 많이 증가하지 못한 것을 발견하였으며, 모병제에 대한 사람들의 관심도는 점차 증가하는 추세를 확인하였다.

주요어 : 감정 분석, 딥러닝, 모병제, 출산율

Abstract A decrease in fertility rates causes problems such as decrease in the working-age population, and has a significant impact on national policies. Currently, the Republic of Korea has a conscription system that imposes military service on all men over the age of 18. However, the transition to the volunteer military system is emerging as a social issue due to the decrease in the fertility rate. In this paper, news articles and comments searched for through the keyword 'volunteer military system' were collected to analyze the social perception of the volunteer military system from 2018, when the fertility rate dropped to less than 1. Some of the collected comments were labeled, and emotional levels were calculated through deep learning models. Through this study, we found that awareness of recruitment system conversion did not increase as the decrease in the fertility rate, and it was confirmed that people's interest is gradually increasing.

Key words : Sentiment Analysis, Deep Learning, Volunteer Military System, Fertility Rates

1. 서론

출산율 감소는 생산가능인구가 감소하고, 인구구조 고령화에 따른 저축률 저하로 자본축적이 줄어들어 경제

성장이 둔화 등의 문제를 발생시킨다. 대한민국의 출산율은 2015년부터 현재까지 매년 감소 추세를 보인다. 2018년에는 출산율이 1 미만으로 감소하면서 사회적 이슈가 되었으며, 2021년 현재까지도 출산율이 1 이상으로

*준회원, 청주대학교 소프트웨어융합학부 학사과정 (제1저자) Received: November 22, 2021 / Revised: January 1, 2022

**준회원, 청주대학교 소프트웨어융합학부 학사과정 (참여저자) Accepted: January 8, 2022

***준회원, 청주대학교 소프트웨어융합학부 학사과정 (참여저자) *Corresponding Author: kafa46@cju.ac.kr

****정회원, 청주대학교 소프트웨어융합학부 교수 (교신저자) Dept. of Software Convergence, Cheongju Univ, Korea

접수일: 2021년 11월 22일, 수정완료일: 2022년 1월 1일

게재확정일: 2022년 1월 8일

오르지 않고 있다. 현재 대한민국 헌법에서는 모든 국민에게 국방의 의무가 있다고 선언하면서 병역법에는 만 18세 이상 남자들에게 국한하여 병역 의무를 부과하고 있다. 하지만 출산율 감소로 인해 병력 유지 문제가 불거지면서, 일각에서는 국방력이 약화 될 것이라는 걱정의 목소리도 나오고 있다. 출산율 감소와 더불어 병력 감소 문제는 현재 대한민국이 해결할 사회적 문제 중 하나이며, 이에 따라 현재 시행하고 있는 징병제를 모병제로 전환하자는 주장도 나오고 있다[1].

2021년 5월 한국 갤럽(www.gallup.co.kr)에서 시행한 여론조사(데일리 오피니언 제449호)를 살펴보면 현재 시행되는 징병제를 유지하자는 의견이 42%, 징병제를 모병제로 전환하자는 주장이 43%로 변한 것을 확인하였다. 특히 2016년 같은 조사와 비교해보았을 때, 징병제를 유지하자는 의견이 6% 정도 감소하였다. 모병제를 전환하자는 의견이 8% 증가한 것으로 보아 최근 5년간 모병제 전환에 대한 찬성 여론이 증가한 것을 확인할 수 있다. 표 1은 한국 갤럽에서 조사한 설문조사 응답 비율이다.

표 1. 설문조사 응답 비율
Table 1. Survey response rates

	2016'09	2021'05
Maintaining the conscription system	48%	42%
Introduction of volunteer military system	35%	43%
No response	17%	15%

해당 여론조사는 전화 조사원 인터뷰 방식을 통해 만 18세 이상 국민 1,003명을 대상으로 진행하였으며, 표본오차는 ±3.1%포인트(95% 신뢰수준)이다. 하지만 여론조사의 표본 집단이 약 1천 명으로 전체 국민 여론을 반영했다고 보기 어려우며, 여론조사의 특성상 질문에 따라 편향된 답변을 유도하기도 한다. 따라서 본 논문에서는 출산율 감소 추세에 따른 모병제 여론 변화의 영향성을 빅데이터 분석 관점에서, 실제 여론조사 결과와 일치하는지 데이터 기반으로 검증하였다.

데이터 기반 영향성의 대표적인 접근법은 기계학습 기반의 감성 분석이다[2,3,4]. 우리는 온라인 뉴스의 댓글 95,879개를 수집하여, 기계학습 기반의 감성 분석을 진행하였다. 데이터 수집 프로그램을 직접 구현하여 ‘모병제’를 키워드로 검색한 뉴스 기사의 댓글을 수집하였다.

수집된 데이터 일부에 대해 Labeling 작업을 진행하였고, 이를 기반으로 딥러닝 학습을 적용하여 감성 분석(Sentiment Analysis)를 진행하였다. 최종적으로 출산율 감소가 모병제 전환 주장에 대한 국민 인식에 어떤 영향을 미쳤는지를 분석하였다.

감성 분석은 직접 설계한 딥러닝 네트워크 기반의 분석 방법과 KoBERT[5] 모델 기반의 분석 방법을 독립적으로 실행하였다. 본 논문에서는 온라인 데이터 수집 및 처리, 딥러닝을 이용한 감성 분석을 이용하여 출산율 저하에 따른 모병제에 대한 감성 분석을 시도하였다. 모병제 관련 언급량, 감성 등을 분석한 결과 출산율 추이에 따른 모병제에 대한 관심도를 데이터 기반으로 분석하고 정량화하여 확인할 수 있었다.

II. 데이터 수집 및 가공

1. 데이터 수집

1) 수집 정보

대중들의 인식을 파악하기 위해 국내 뉴스 기사 댓글을 수집하였다. 데이터 수집 공간을 확정하기 위해 ‘인터넷 트렌드(www.internettrend.co.kr)’를 활용하여 2018년 6월 1일부터 2021년 6월 1일까지 검색 엔진별 순위를 확인하였다. 국내 포털사이트 점유율이 설정한 수집 기간 중 1위는 네이버(60.5%), 2위는 구글(31.4%), 그리고 3위는 다음(6.08%)으로 나타났다. 2위인 구글은 뉴스 댓글을 수집하기에 부적절한 웹 페이지로 판단하여 수집 대상에서 제외하였다. 수집 키워드는 ‘모병제’로 설정하였다. 출산율이 1 미만으로 떨어진 2018년부터의 데이터 모니터링을 위하여, 2021년 6월 1일을 기준으로 최근 3년 동안을 감성 모니터링 기간으로 정하였고, 해당 기간의 뉴스 기사 댓글을 수집하였다.

2) 수집 방법

본 연구에서는 데이터 자동 수집 로봇(크롤러, Crawler)을 직접 구현하여 다음과 네이버 뉴스로부터 설정한 수집 기간과 키워드에 해당하는 모든 뉴스 기사 댓글을 수집하였다. 크롤러 프로그램은 플랫폼마다 뉴스 제공 형태가 상이하여 네이버 크롤러, 다음 크롤러 2개로 구현하였지만, 수집 알고리즘은 같으며, 전체적인 처리 과정을 표현한 알고리즘은 그림 1과 같다.

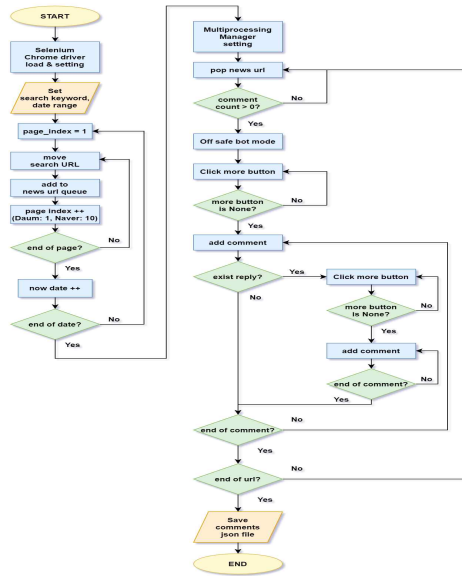


그림 1. 데이터 수집 알고리즘
 Figure 1. Data collection algorithm

수집 기간인 '18.06.01. ~ '21.06.01. 사이의(1,096일간) “모병제” 키워드로 검색한 기사의 링크를 먼저 수집하였고, 이후 댓글 수집을 위해 각 링크들에 해당하는 페이지에 접속하여 댓글이 존재하는 뉴스 기사에 대해 댓글 수집을 진행했다. 다음과 네이버 뉴스 기본 설정을 “모든 댓글 표시”로 변경하여 모든 댓글이 표시되도록 크롤러를 구현하였다. 비속어 등이 포함된 댓글을 제거하는 “세이프봇” 모드도 비활성화하여 모든 댓글을 수집하였고, 수집 결과는 json 형식 파일로 각각 저장하였다.

3) 수집 결과

수집 기간 '18.06.01. ~ '21.06.01. 사이의(1,096일간) “모병제”를 키워드로 다음 뉴스와 네이버 뉴스에서 수집된 댓글의 수집 결과는 표 2와 같다.

표 2. 데이터 수집 결과
 Table 2. Result of data collection

Collection Period	Article	Comment
2018 (6~12)	108	15,684
2019 (1~12)	577	35,150
2020 (1~12)	238	16,644
2021 (1~6)	414	28,401
Total	1,337	95,879

2. 데이터 가공

1) 데이터 전처리

데이터 전처리(data pre-processing)는 데이터를 분석 및 처리에 적합한 형태로 만드는 과정으로, 컴퓨터가 처리할 수 있는 형태로 변형하고 불필요한 정보를 제거하여 인공지능의 예측 정확도를 향상하는 것이 목적이다. 본 연구에서 수집된 모든 데이터는 전처리 과정을 거쳤다. 한글 데이터는 조사에 따른 다양한 의미 변화가 가능하기 때문에 단어분절(tokenization) 작업이 필요하다. 이를 위해 KoNLPy의 Mecab을 활용하였다. KoNLPy에는 Komoran, Kkma, Hannanum, Okt 등의 형태소 분석 라이브러리가 존재한다. 본 연구의 형태소 분석을 위해 가장 성능이 우수하고 형태소 분석 속도가 빠르다고 알려진 Mecab 라이브러리를 선정하였다. 형태소 분석이 완료된 댓글에서 숫자(SN), 조사(JKS, JKC, JKG, JKO, JKB, JKV, JKQ, JC, JX), 기호(SE, SSO, SSC, SC, SY), 한자(SH)를 제거하여 데이터 전처리 작업을 하였다. 표 3에서 Raw data는 전처리 이전이며, Preprocessed data는 Mecab 라이브러리를 활용하여 전처리를 진행한 결과이다.

표 3. 데이터 전처리 예시
 Table 3. Example of data preprocessing

Raw data
통일이 된후에 모병제를 해야 한다
Preprocessed data
통일 된 후 모병제 해야 한다

2) 감성 레이블링

수집한 데이터에는 감성 레이블이 존재하지 않기 때문에, 댓글 데이터의 일부를 직접 레이블링 하였다. 레이블링 작업은 모병제에 긍정적인 경향을 보이는 댓글은 1, 부정적인 경향을 보이는 댓글은 0으로 레이블링 하였다. 레이블링 작업 시 개인마다 긍·부정을 판단하는 기준이 상이하야 편향이 존재할 수 있다. 이 점을 보완하기 위하여 2명의 인원이 추가 검토를 수행하였다. 의견이 다른 댓글에 대해서는 과반수인 2명이 선택한 레이블 결과를 따르도록 하였다. 이를 통해 특정 인원의 편향이 제거되도록 노력하였다. 총 9,820개의 댓글에 감성 레이블링을 진행하고 훈련(Train) 데이터를 생성하였다.

III. 감성 분석

1. 딥러닝 모델 구현

본 연구에서는 딥러닝 모델을 구축하기 위해서 Python package 중 하나인 Keras를 이용하였다. 입력층과 출력층 사이에 11개의 은닉층을 가진 딥러닝 모델을 구성하였다(그림 2 참조).

첫 번째 은닉층은 Keras에서 제공하는 내장 함수 중 문자열을 벡터화해주는 함수인 TextVectorization을 활용하였다. 이때, TextVectorization의 인자로서 전체 단어 사전의 크기를 의미하는 max_tokens는 100,000, 출력 모드를 결정하는 output_mode는 정수(int)형, 출력 벡터의 크기를 나타내는 output_sequence_length는 64로 구성하였다.

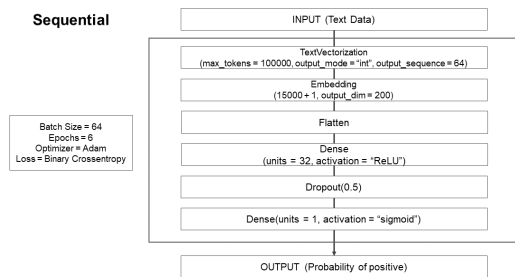


그림 2. 딥러닝 모델 구조
Figure. 2. Deep learning model structure

훈련 데이터를 TextVectorization 함수에 적용하여 1 번째 출력 벡터를 생성하였다. 두 번째 은닉층은 벡터화된 문자열인 양의 정수(색인)를 고정된 크기의 밀집 벡터로 전환해주는 Embedding 함수를 활용하였다. Embedding 함수의 인자 중 어휘 목록의 크기 즉, 최대 정수 색인+1이 되는 input_dim은 TextVectorization의 인자 중 하나인 max_tokens 값에 1을 더해준 값으로 설정하였다. 출력값의 밀집 벡터 차원을 결정하는 output_dim은 200으로 설정하였다. 세 번째 은닉층은 입력한 밀집 벡터를 평평하게 해주어 배열로 전환해주는 Flatten 내장 함수를 활용하였다. 네 번째부터 열 번째 은닉층은 처음 256개의 뉴런(Neuron)에서 은닉층이 추가될 때마다 뉴런이 기존 뉴런 개수의 절반으로 감소하는 형태로 설계하였으며, 모든 은닉층에서의 활성화 함수(Activation Function)는 ReLU(Rectified Linear Unit)로 설정하였다. 이때, 훈련 과정에서의 과적합(Overfitting)을 방지하기 위해 은닉층에 Dropout을 적용하였다. Dropout은 훈련 과정에서 신경망 노드를 일시적으로 무작위하게 제거하는 방법이며, 본 연구에서는 Dropout의 비율을 0.5로 설정하였다.

출력층은 1개의 노드를 가지고, 활성화 함수를 sigmoid로 설정하였다. 하나의 댓글 데이터 당 하나의 출력값만 존재해야 하므로 출력 노드는 1개이고, 0부터 1 사이의 소수 출력값을 가져야 하므로 sigmoid 함수를 사용하였다. 이때 출력값은 1에 가까울수록 긍정적인 경향을 나타내며, 0에 가까울수록 부정적인 경향을 나타낸다. 신경망이 학습할 수 있도록 해주는 지표가 바로 손실 함수(Loss Function)이다.

본 연구에서는 두 개의 Binary 값 중 하나를 선택하는 것이 목적이며, 데이터의 레이블이 독립적이기 때문에 Binary Cross Entropy를 손실 함수로 설정하였다. 신경망 학습은 이러한 손실 함수의 지표를 토대로 손실 함수값을 최대한 낮추는 최적의 매개변수를 찾아야 한다. 이러한 과정을 최적화(Optimization)라고 부른다. 최적화를 위한 옵티마이저(Optimizer)는 Keras 모델을 컴파일하기 위해 필요한 변수 중 하나이다. 본 연구에서는 신경망 학습을 하면서 경사 하강법(Gradient Descent)의 일부인 Adam을 옵티마이저로 설정하였다. Adam은 학습률을 줄여나가고 학습 관성 속도를 계산하여 학습의 갱신 강도를 조정하는 알고리즘이다.

본 연구에서 직접 감성 레이블링한 데이터는 총 9,820개로 전체 데이터의 약 10% 이다. 해당 데이터셋을 활용하여 Train과 Test 데이터셋을 8대 2의 비율로 구성하였다. 신경망 학습 결과, 훈련 데이터의 정확도(Accuracy)는 약 97%, 손실(Loss) 값은 약 0.047의 성능을 보였다. 테스트 데이터의 정확도는 약 95%이고, 손실 값은 약 0.142로 나타났다.

생성된 딥러닝 모델을 잔여 댓글 데이터에 적용하였다. 총 95,879개의 댓글에 대해 훈련 데이터와 같은 방식으로 데이터 전처리를 진행한 다음, 훈련된 모델을 통해 각 댓글의 감성 수치를 예측하였다. 예측된 감성 수치는 최종 출력 노드의 활성화 함수인 sigmoid를 통과시켜 0.0 이상 1.0 이하의 실수로 출력하였다.

2. KoBERT 모델 활용한 분석

BERT 모델은 대규모 데이터셋을 사전 학습하여 간단한 Fine-tuning을 통해 여러 가지 자연어 처리 문제를 해결할 수 있도록 고안되었다[3]. BERT는 transformer의 인코더 부분을 활용한 모델이며, 최근 자연어 처리의 여러 태스크에서 가장 성능이 높은 아키텍처로 알려져 있다. 본 연구에서는 SKTBrain에서 개발한 한국어 BERT

모델인 KoBERT의 PyTorch API를 활용하였다. 이때, Train data와 Test data를 8:2으로 구분하였다. 데이터 최대 길이 max_len은 64, batch size는 32, epochs는 5, learning_rate는 $5e-5(10^{-5})$, dropout의 비율은 0.4로 설정하였다. 학습 결과, 훈련 데이터의 정확도는 약 98%, 손실 값은 약 0.018의 성능을 보였다. 테스트 데이터의 정확도는 약 94%로 나타났다. 생성된 모델을 3.1과 같은 방식으로 전체 댓글에 적용하였다.

3. 회귀 분석을 통한 감성 값 산출

모든 댓글 데이터에 대해 딥러닝 모델을 활용하여 감성 레이블링을 진행하였다. 최초 날짜당 감성 값은 가중 평균(Weight Average)을 이용해 도출하였다. 날짜당 기사가 k 개 있다고 가정할 때, 기사 인덱스 i 의 총 댓글 수는 $count(i)$, 기사 평균 감성 값은 $avg(i)$ 라고 하겠다. 계산의 편의를 위해 해당 날짜 모든 기사의 총 댓글 수의 합을 M 으로 표현한다. 날짜당 가중 평균 W 를 구하는 방법은 식 (1)과 같다.

$$W = \sum_{i=1}^k (count(i)/M) \times avg(i) \quad (1)$$

식 (1)은 날짜별 댓글 수의 편향이 발생할 수 있다. 이를 보정하기 위해 본 논문에서는 날짜별로 가중 평균을 구한 다음, 전체 기간에서 해당 날짜의 댓글 수를 고려하기 위해 가중 평균값에 새로운 공식을 적용했다. 이 공식은 IMDb에서 투표수를 고려하여 영화 순위를 측정하는 방식을 응용하였다[6].

$$X = \frac{N}{L+N} \times W + \frac{L}{L+N} \times e \quad (2)$$

날짜당 총 댓글 수를 N , 감성 수치가 신뢰를 얻기 위한 날짜당 최소 댓글 수를 L 이라고 한다. 본 연구에서 L 은 전체 기사에 포함된 댓글 수의 평균값인 108개 이하의 댓글을 보정하기 위해 108로 설정하였다. 전체 기간의 감성 평균을 e , 그리고 날짜당 가중 평균값을 식 (1)에서 구한 W 를 사용한다. 이때 전체 기간에서 해당 날짜의 총 댓글 수를 고려한 날짜별 감성 평균값 X 는 식 (2)와 같다.

4. 긍정 비율을 통한 감성 값 산출

딥러닝 모델을 활용하여 감성 레이블링을 진행할 때, 0.5 미만의 예측값을 가진 댓글은 0으로, 0.5 이상의 예측값을 가진 댓글은 1로 분류(판정)하였다. 특정 날짜의 감성 값을 산출하기 위해 별도의 수학 공식을 적용하지 않고, 모든 댓글 중에서 긍정으로 분류된 댓글(예측값 0.5 이상)의 비율을 계산하였다.

IV. 연구 결과 분석

1. 회귀 분석을 이용한 결과

3.3에서 제시한 회귀 분석 방법을 통해 두 가지 모델에서 예측된 감성 값을 계산하여 특정 날짜의 감성 값을 산출하였다.

1) Keras 모델을 활용한 결과

본 연구를 위해 직접 설계 및 구현 Keras framework 기반의 딥러닝 모델을 활용하여 전체 댓글에 대한 감성 값을 예측한 결과를 분석하였다. 분석 결과 그래프에서 Y축은 0.0 이상 1.0 이하의 실숫값으로 1.0에 가까울수록 긍정적인 경향을 보이고, 0.0에 가까울수록 부정적인 경향을 보인다. 이때 감성값이 0.3 이상인 데이터가 없는 점을 고려하여, 그래프에서 Y축의 범위는 0부터 0.3로 설정하였다. X축은 2018년 6월 1일부터 2021년 6월 1일까지 기사와 댓글이 하나 이상 수집된 날짜를 오름차순으로 나열하였으며, 년과 월 정보만 숫자로 표시하였다. 키워드별로 수집한 기간에 따라 나타난 감성 결과는 그림 3에 제시하였다.

2) KoBERT 모델을 활용한 결과

KoBERT 모델의 PyTorch API를 활용하여 학습한 딥러닝 모델을 활용하여 전체 댓글에 대한 감성 값을 예측하였다. KoBERT 결과는 그림 4에 제시하였다.

KoBERT 모델에서는 감성값이 0.4 이상인 데이터가 없는 점을 고려하여, Y축의 범위는 0부터 0.4로 설정하였다. X축 정보는 4.1.1에 설명한 것과 동일하다.

두 모델을 통해 3년 동안의 감성 분석을 예측한 그림 3,4의 형태가 비슷한 것으로 보아, 모델 성능이 우수할 것으로 볼 수 있다. 하지만 두 그래프를 비교해볼 때, 의미 있는 변화를 관찰하기 어렵다.

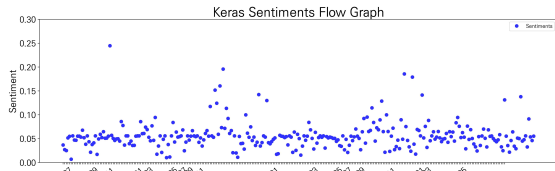


그림 3. Keras 모델 기반의 회귀 분석 그래프
Figure 3. Keras model-based regression graph

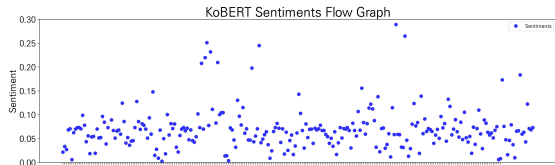


그림 4. KoBERT 모델 기반의 회귀 분석 그래프
Figure 4. KoBERT model-based regression graph

2. 긍정 비율을 이용한 결과

회귀 분석을 통해 감성 값을 산출한 결과로 감성 분석의 결과를 단정하기 어려워 감성 분류 모델을 긍정 비율을 계산하는 방식으로 변경하여 결과를 다시 확인하였다.

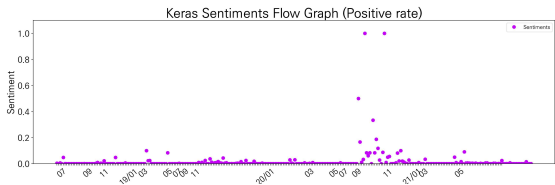


그림 5. Keras 모델 기반 긍정률 그래프
Figure 5. Keras model-based positive rate graph

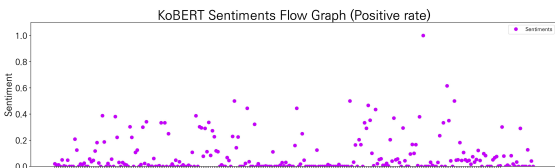


그림 6. KoBERT 모델 기반 긍정률 그래프
Figure 6. KoBERT model-based positive rate graph

두 모델에서 예측된 감성 값을 활용하여 긍정 비율을 계산하는 방식은 3.4에 제시하였다. 4.1과 같은 방식으로 Keras 모델과 KoBERT 모델을 활용한 결과를 그래프로 생성하였으며, 각각 그림 5,6에 제시하였다. 해당 그래프에서 X축은 4.1.1의 설명과 같으며, Y축은 산출된 감성 값으로, 두 모델 모두 0.0부터 1.1 범위를 가진다. 두 모델의 감성 산출 결과에서 긍정률을 나타낸

그래프는 전체 기간에서 지속하여 낮은 수치를 보인다. 두 그래프에서 상대적으로 긍정률이 높아지는 구간을 추출하였고, 그림 5에서 2020년 7월 8일과 2020년 9월 17일, 그림 6에서 2020년 12월 14일로 나타났다.

하지만 두 모델에서 긍정률이 높아지는 3일에서 모두 댓글 수가 3개 이하로 나타났으며, 표본 집단이 작아서 긍정률의 신뢰성을 보장할 수 없다. 따라서 전체 데이터에서 특정 날짜의 총 댓글 수가 5개 미만인 경우의 날짜는 제외하고 다시 그래프를 생성하였다(그림 7, 8 참조).

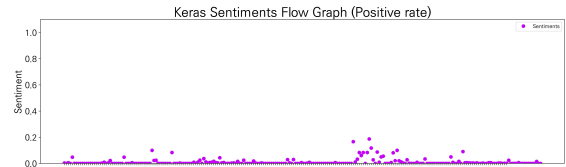


그림 7. 댓글 5개 미만의 Keras 모델 기반 긍정률 그래프
Figure 7. Keras model-based positive rate graph with more than 5 comments

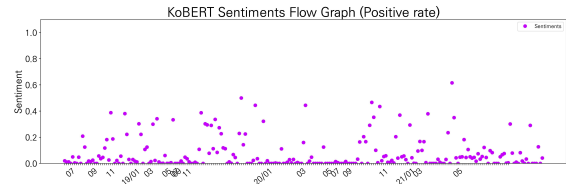


그림 8. 댓글 5개 미만의 KoBERT 모델 기반 긍정률 그래프
Figure 8. KoBERT model-based positive rate graph with more than 5 comments

날짜당 댓글 수가 5개 이상인 날짜만 표시한 그래프에서 긍정률이 상대적으로 높아지는 구간이 발생하지 않은 것을 확인할 수 있다.

Keras 모델과 KoBERT 모델 기반의 감성 예측 수치를 회귀 분석 방법과 긍정 비율 분석 방법을 통해 확인한 예측 감성 흐름이 3년간 전체 기간에서 꾸준히 낮은 수치를 보인다. 출산율이 감소함에 따라 모병제 전환에 긍정적인 변화가 관찰되지 않았으며, 한국 갤럽에서 조사한 여론조사의 결과와 다소 다른 결과를 보인다. 여론조사에서는 모병제 전환에 43%의 찬성 여론이 나타났다. 그러나 실제 데이터 기반의 분석 결과에서는 전체 기간에서 모병제 전환에 부정적 의견이 주를 이루고 있는 것을 확인하였다.

3. 기간별 평균 및 표준편차

2018년 6월부터 2021년 6월까지 1년간 3개의 구간으로 구분하여 감성값의 평균과 표준편차를 분석하였다. 그림 9는 Keras 모델 기반의 회귀 분석 방법을 적용한 감성 산출 값의 통계이다.

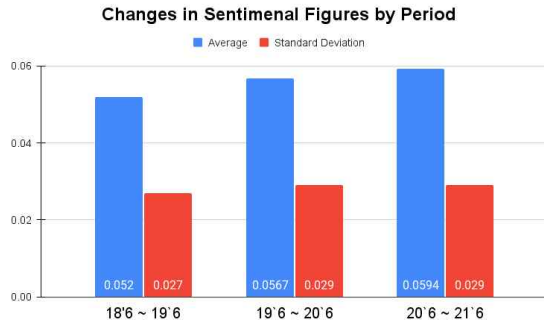


그림 9. 기간별 평균 및 표준편차 그래프

Fig. 9. Graphs of mean and standard deviations by period

세 구간에서 평균은 각각 0.052, 0.0567, 0.0594이며, 표준편차는 각각 0.027, 0.029, 0.029로 나타났다. 의미 있는 차이를 보이지 않지만, 시간의 흐름에 따라 평균과 표준편차가 증가함에 따라 사람들의 여론이 다양한 의견으로 확대됨을 확인하였다.

V. 결 론

본 연구는 국내 출산율 변화에 따라 현 징집 제도인 징병제를 모병제로 전환하는 인식의 변화를 알아보았다. 2021년 6월부터 출산율이 1 이하로 떨어지는 시점인 2018년까지 최근 3년간의 데이터를 수집하여 딥러닝을 통해 분석하였다.

연구 결과 시간의 흐름에 따라 출산율은 저하되었지만, 모병제에 대한 인식 변화는 크게 변화되지 않았다. 또한 모병제에 부정적인 인식이 지속되고 있는 것을 확인하였다. 한국 갤럽의 여론조사에서 모병제 전환에 찬성 여론이 43%인 것에 비해 다소 상이한 결과가 나타남을 데이터 기반으로 확인하였다. 따라서 표본 집단이 작은 수준에서, 전화를 통한 여론조사 방식이 실제 여론을 충분히 반영하지 못한다는 한계점을 발견하였다. 하지만 모병제 키워드 관련 댓글 수가 증가하였고, 감성 평균이 증가한 것을 보았을 때 사람들의 인식이 점차 변화하고 있다는 점을 확인하였다.

References

- [1] M. Kim, "A Study on the Transition of Military Recruitment System in Korea -Focusing on the Transition of the Taiwan Recruiting System," Asia Culture Academy of Incorporated Association, Vol 21, pp.883-898, 2021. <http://dx.doi.org/10.22143/HSS21.12.1.62>
- [2] S. Ahn, S. Ryu, and S. Hong, "A sentiment analysis model for small-scale unstructured policy data using transfer learning," Journal of the Korean Data And Information Science Society, Vol. 31, No. 2, pp. 405-414, 2020. <https://doi.org/10.7465/jkdi.2020.31.2.405>
- [3] S. Park and K. Lee, "Effective Korean sentiment labeling method using word embedding and semi-supervised learning," Journal of Korean Institute of Intelligent Systems, Vol. 28, No. 2, pp. 185-191, 2018. <https://doi.org/10.5391/JKIS.2018.28.2.185>
- [4] S. H. HA and T. H. ROH, "Sentiment Analysis for Public Opinion in the Social Network Service," The journal of the convergence on culture technology, Vol. 6, No. 1, pp. 111 - 120, Feb. 2020. <https://doi.org/10.17703/JCCT.2020.6.1.11>
- [5] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," arXiv preprint arXiv:1810.04805, 2018. <https://doi.org/10.18653/v1/N19-1423>
- [6] M. Sahay. "How does IMDb's rating system work?" Quora. <https://www.quora.com/How-does-IMDb's-rating-system-work> (accessed Jun 26, 2021).