

웹크롤링과 텍스트마이닝을 활용한 국군간호사관학교와 간호장교 인식도 조사

이한인¹, 차진호^{2*}

¹국방부, ²국방기술진흥연구소

A Study on the Recognition of Korea Armed Forces Nursing Academy and Military Nursing Officers through Web Crawling and Text Mining

Hanin Lee¹, Jinho Cha^{2*}

¹Ministry of Defense

²Korea Research Institute for defense Technology planning and advancement

요약 국군간호사관학교(국간사)는 한 때 폐교될 위기에 처한 적이 있다. 기업과 마찬가지로 군대내 부대나 조직도 정체성을 스스로 진단하고 그 필요성을 입증해 나가지 않으면 그 존재를 장담할 수 없을지 모른다. 국간사의 폐교 위기 전후 국간사와 간호장교의 역할에 대해 군 내부적으로 진단하는 연구가 있었지만 군 외부에서 바라보는 시각에 대한 연구가 없었기 때문에 정확한 진단을 하는 것이 제한되었다. 실제로 군 외부에서 국민들이 바라보는 시각에 대한 분석은 군의 신뢰도를 분석하는데 밀접한 관련이 있고 이러한 군의 신뢰도는 군의 무형적인 전투력인 사기와 연관되고 전투나 전쟁의 승패에 큰 영향을 미칠 수 있다. 본 연구는 국민들이 국간사와 간호장교를 대상으로 온라인상에서 2011년부터 2020년까지 개진한 방대한 양의 데이터를 웹크롤링을 통해 수집하고 텍스트마이닝 기법으로 인식도를 분석하였다. 분석 시 사용한 K-Means 클러스터링과 Doc2Vec 기법은 의견으로 제시된 단어를 벡터화하고 주변 단어들과의 관련성을 유지하면서 군집 형성을 수행하는 기법을 적용함으로써 군집별 의미론적 분석의 신뢰성을 보장하고 있다. 본 연구 기법은 국간사와 간호장교를 군 외부에서 바라보는 인식을 텍스트마이닝의 토픽모델링 기법을 적용한 분석이라는 의미를 가지며 연구 결과 군 외부 인식이 긍정적임을 확인할 수 있다.

Abstract The Korean Armed Forces Nursing Academy (KAFNA) once faced the possibility of being shut down. Just like enterprises, The survival of military units or organizations within the military cannot be guaranteed if they do not diagnose their identity and prove its necessity. On the other hand, an accurate diagnosis is limited because there has been no research on the roles of KAFNA and nursing officers from an external perspective before and after the crisis of the KAFNA closure. Indeed, analyzing external perspectives is closely related to analyzing the credibility of the military, and such credibility can have a significant effect on the outcome of battles or wars. This study used web-crawling and text-mining techniques to analyze a vast amount of data on KAFNA and nursing officers posted online from 2011 to 2020. The K-Means clustering and Doc2Vec techniques used in the analysis vectorize the words presented as opinions and perform clustering while maintaining the relevance with surrounding words, ensuring the reliability of semantic analysis by cluster. This research technique is the first text-mining technique to analyze the perception of KAFNA and nursing officers from an external perspective, and the study results show that the external perception is positive.

Keywords : Nursing, Military, Web Crawling, Text Mining, K-Means, Doc2Vec

*Corresponding Author : Jinho Cha(Korea Research Institute for defense Technology planning and advancement)
email: jinchocha7kma@krit.re.kr/jinchocha7@gmail.com

Received March 22, 2023

Revised April 26, 2023

Accepted May 12, 2023

Published May 31, 2023

1. 서론

1.1 연구의 필요성

코로나바이러스 감염증-19(이하 코로나19) 발생초기 민간 의료지원을 위해 투입된 군 인력 중 국군간호사관학교(이하 국간사) 출신 간호장교가 임관식을 앞당긴 후 전원 투입되었다[1]. 위 사례는 TV 등 매스컴을 통해 몇 달간 거의 매일 방송이 되었고 이로 인해 국간사 생도와 간호장교들을 바라보는 국민의 인식이 개선되었다고 생각할 수 있다.

하지만 20여년 전 국간사는 2년간 신입생도들을 선발하지 않았고 폐교가 될 위기에 있었다. 당시 국간사가 필요하냐는 외부의 질문에 군내부에서 합리적인 답변을 내놓지 못했다. 국민이 바라보는 국간사의 인식을 사전에 조사하고 이에 대응하는 방안을 체계적으로 가졌다면 당시의 논의로 인한 폐교 위기 상황이 발생하지 않았을 수도 있다. 4차 산업혁명으로 인한 과학화된 데이터 분석 기술의 발전으로 최근에는 국민의 인식을 직접 설문하지 않고 온라인상에 기술되어 있는 다양한 의견을 수집 및 조사할 수 있는 환경이 마련되었다. 이를 통해 국민들이 군대를 바라보는 인식을 보다 솔직하고 객관적으로 진단하고 확인해 볼 수 있게 되었을 뿐만 아니라 대국민 이미지 개선을 위한 신속한 대응이 가능해졌다.

본 연구에서 사용한 웹크롤링(Web crawling)이라고 불리는 방법을 통해 온라인 웹사이트에 있는 텍스트나 이미지와 같은 비정형 데이터를 획득할 수 있는 방법은 최근 활발하게 응용되고 있다[2]. 웹크롤링 기법은 국간사와 간호장교에 대해 웹상에서 개진한 다양한 의견들을 일반적인 설문조사와 다르게 시간과 장소에 제약이 없이 수집할 수 있게 한다. 그리고 이렇게 얻은 텍스트와 같은 비정형데이터는 텍스트마이닝(Text mining)이라고 불리는 기법을 통해 그 의미와 정보를 과학적으로 분석할 수 있다[3]. 텍스트마이닝을 웹상에 있는 데이터에 적용하는 웹마이닝(Web mining)은 주로 인터넷에 작성된 수많은 의견이나 글을 수집한 후 의미론적 분석(Semantic analysis) 방법론을 사용하여, 특정 주제에 대한 군집별 세부적인 의견을 도출해내는 과정으로 주로 진행되었다[4]. Ril 등의 연구[5]는 트위터에 올라온 의견에 포함된 의미가 문서-주제 확률분포에 근거하여 일정 개수의 주제로 분류하고, 각 주제별로 의견을 분석하여 정치적인 성향을 분석하는 모델을 설계하였다. 그러나 의미론적 분석을 통한 집단의 의견을 도출하는 연구 방법은 의미 분석에 사용되는 요소들의 단조로움으로 인해

정확성이 낮아 방대한 양의 다양한 의견을 분석하기에는 어려움이 있었다[4]. 그 결과 군집별 의견을 분석함에 있어 개선된 분석결과를 도출하기 위해 기존의 의미론적 분석 방법에 새로운 연구방법을 결합하는 연구가 진행되었다. Zhang 등의 연구[6]에서는 문서의 단어간 동시발생 행렬을 의미 분석에 접목시켜, 단어간의 구조를 그래프화하여 의미 분석을 용이하게 해주는 주요 키워드를 도출하였으며, Yan 등의 연구[7]에서는 단어-문서 역빈도(TF-IDF, Term Frequency-Inverse Document Frequency) 점수를 이용해 의견 분석의 정확성을 높였다. 또한 Lin 등의 연구[8]은 비지도학습의 대표적인 군집화 방법 중에 하나인 K-means 클러스터링 단어를 기반으로 적용해서 전체 문서의 주제를 분류하였다.

한편 기존의 의미론적 분석 방법에서 벗어나 신경망구조(Neural network) 기반의 문서 임베딩인 Doc2Vec을 사용하는 연구도 이루어지고 있다[9]. Staykovski 등의 연구[10]은 Doc2Vec과 TF-IDF 점수를 통해 뉴스데이터의 주제를 분석하여 문서를 분류하는 연구를 진행하였으며, Rădulescu 등의 연구[11]은 기존 문서를 그래프로 구조화하여 여러 집단으로 분류한 후 Doc2Vec과 K-means 클러스터링을 통해 군집화에 대한 정확도를 향상시켰다. 이와 같이 텍스트마이닝의 연구분야에서 특정 주제와 관련된 여러 인식을 분석하고 그에 따른 군집화된 집단을 도출해내는 연구는 활발히 진행중이며, 단어간의 의미론적 분석의 차원에서 단어 혹은 문장기반의 임베딩 형식으로 발전되었다.

한편 국간사 생도들과 간호장교를 바라보는 군 외부에서의 인식도 조사에 대한 기존의 연구는 거의 진행되지 않았다. Chu 등의 연구[12]는 군병원 의료에 대한 수준을 연구하기 위해 군의관, 간호장교, 환자를 대상으로 진행하였고, Na 등의 연구[13]은 군병원에 입원한 환자가 바라보는 간호장교에 대한 이미지에 대해 수행하였으며, Woo 등의 연구[14]는 국간사 생도 380여명을 대상으로 군에 대한 소명인식을 분석하기 위해 국간사 생도 380여명을 대상으로 진행하였다. 위 3편의 논문들은 군내부 특정인원에 대한 수개월의 단기간 동안 설문조사를 통해 실시한 연구들이다. 2020년 2월부터 3월까지 대구지역의 코로나19 집단감염으로 국간사 신입소위들 전원이 대구 소재 병원으로 투입되어 근무한 경험을 바탕으로 작성한 연구도 군 외부에서 바라보는 인식에 대한 논문이 아니다[15].

이에 본 연구에서는 최근 10년간의 국간사 생도와 간호장교에 관련된 뉴스 기사에 국민들이 의견을 개진한

댓글을 웹상에서 크롤링을 통해 수집 후 분석하여 군 자체에서 바라보는 인식도가 아닌 군외부에서 군을 바라보는 대국민 인식도에 대한 결과를 객관적으로 제공하고자 한다.

1.2 연구의 목적

본 연구의 목적은 2011년부터 2020년까지 국간사와 간호장교 관련 네이버 뉴스 기사에서 수집한 댓글을 텍스트마이닝 기법으로 분석하여 군 외부에서 어떠한 인식을 가지고 있는지를 분석하기 위함이며, 구체적인 목적은 다음과 같다.

첫째, 군 내부 인원이나 관련자에 의한 설문이나 통계 조사가 아닌 군 외부에서 바라보는 국간사와 간호장교에 대한 인식도를 폭넓게 조사하여 그 결과를 가시화하여 제공한다.

둘째, 객관성을 보장하기 위해 수개월(단기)이 아닌 수년간(중기)의 데이터를 수집하고 최신 텍스트마이닝 기법을 적용하여 과학적으로 분석한다.

셋째, 분석된 결과를 통해 군 외부 인식 정도를 확인하고 개선을 위한 수단으로 사용하거나 향후 장기간의 데이터를 추가 수집하여 연구할 수 있는 이론적 근거를 제공한다.

2. 본론

2.1 연구 설계

본 연구는 텍스트마이닝을 이용하여 국간사와 간호장교에 대한 대국민 인식도를 조사하기 위한 방법론적 연구이다.

2.2 연구 대상자

본 연구의 대상자는 2011년부터 2020년 간 국간사와 간호장교에 관련된 네이버 뉴스에 의견을 개진한 온라인상의 국민들의 의견이며 중복을 제거하고 분석에 사용된 전체 데이터는 11,433건이다.

2.3 연구수행 절차

본 연구는 간호장교와 관련한 인터넷 기사를 분석해 어떠한 집단으로 간호장교에 대한 의견이 나뉘는지 분석하고, 각 집단 별로 단어를 추출하여 의미를 발견하고자 하였다. 전체적인 연구의 구조는 Fig. 1에서 보는 바

와 같다. 첫 번째 단계에서는 인터넷 비정형 데이터인 간호장교 관련 뉴스 기사를 정형적인 벡터로 변환한다. 파이썬 크롤러를 이용해서 네이버에 '간호장교'라는 키워드가 입력된 뉴스 기사 전체를 수집하였다.

수집한 전체 기사로부터 텍스트 전처리를 진행하여, 불용어를 제거하였으며 모든 단어를 토큰화하였다. 토큰화된 모든 데이터는 신경망구조 기반의 임베딩 알고리즘, Doc2Vec을 통해 벡터화하였다[9].

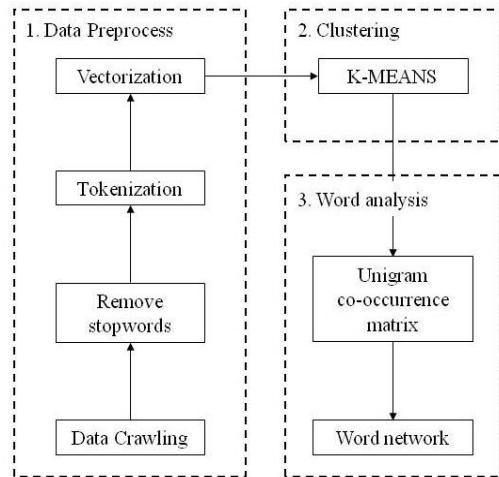


Fig. 1. Research framework

두 번째 단계에서는 벡터화된 데이터에 비지도학습인 K-means 클러스터링을 통해 군집화하였다. 군집화를 통해 나누어진 집단별로 어떠한 단어가 주로 나타나는지 분석하기 위해 '간호장교'라는 단어를 중심으로 동시발생행렬을 만들어 시각화하였다. 이와 같은 전체적인 과정을 통해 간호장교에 대한 여러 의견이 집단별로 구분되며, 각각의 집단이 어떠한 의견으로 식별되는지 연구하였다.

2.4 연구 도구

본 연구는 간호장교와 관련한 인터넷 기사를 분석해 간호장교에 대한 의견을 그룹별로 분석하기 위해 Doc2Vec이라는 문서기반으로 벡터를 변환하는 기법과 분류된 벡터들을 유사한 그룹별로 나누는 K-means 군집화하는 방법을 사용한다.

2.4.1 문서기반 벡터

Doc2Vec은 다양한 길이의 비정형 텍스트를 같은 크

기의 연속적인 값을 가진 벡터로 변환시키는 방법론이다. 기존의 말뭉치(Bag-of-words) 방식의 텍스트 표현은 단순한 단어의 빈도수에 기반하여 의미를 분석하는 방법인데, 이는 문장이 가진 정확한 의미를 재현하는데 한계가 있다. 이에 반해 Doc2Vec으로 표현이 되는 문서기반의 벡터는 각각의 단어가 가진 주변 단어들과의 관련성을 유지하면서 학습되어지므로, 문장의 의미를 유지한 채로 투영(Projection)된다. Rădulescu 등의 연구 [11]은 Doc2Vec이 문서 단위의 임베딩으로 단어가 아닌 문서가 가진 의미인 문맥 자체를 벡터로 재현 가능하도록 한다. 이러한 문서단위의 임베딩 방식은 기존의 확률을 이용한 문서를 분류하는 연구에 비해 우수한 분석 성능을 가진다[9].

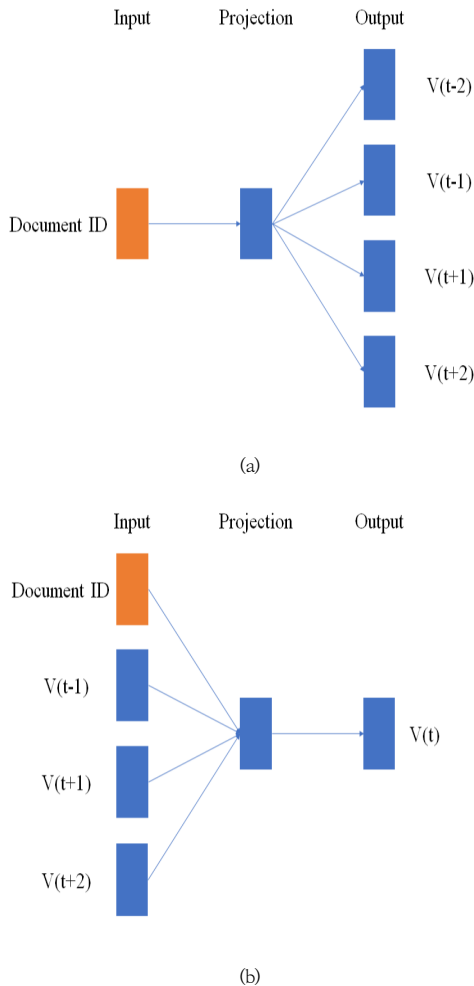


Fig. 2. DM and DBOW structures of Doc2Vec
 (a) Distributed memory(DM)
 (b) Distributed bag of words(DBOW)

이러한 Doc2Vec의 학습방법은 Fig. 2에서 보는 바와 같이 (a) Distributed Memory(DM) 모델과 (b) Distributed Bag of Words(DBOW) 모델 두 가지로 나뉘어진다. DM모델은 각각의 문서 ID가 기존 Word2Vec의 Continuous bag of words 방법론[16]에 추가된 모형으로, Window 사이즈 내의 빈 공간의 단어를 주변 단어를 통해 예측한다. 이를 통해 학습에 포함된 문서의 ID가 단어와 마찬가지로 벡터화되어 전체 문서의 의미가 투영된 행렬을 얻게 된다. DBOW 모델은 문서의 ID가 주어졌을 때 해당하는 단어를 예측하는 방법이다. 단어벡터를 저장하는 방식인 DM과는 다르게 전체 단어를 학습하지 않는 방식으로, 각 문서 ID마다 다른 단어 빈출 확률을 통해 모델이 학습된다. 일반적으로 DBOW 모델보다 DM모델의 성능이 좋다고 알려져 있다[9].

2.4.2 K-means 클러스터링

K-means 클러스터링 알고리즘은 사용자가 지정한 K개의 중심점(Centroids)을 기반으로 벡터데이터 내에서 중심점에 가까운 데이터를 하나의 군집으로 묶는 알고리즘이다[17]. 첫 번째로 군집의 개수 K개를 선정하면 데이터가 분포된 공간상에 임의의 지점 K개를 지정한다. 두 번째로 임의의 중심 K개로부터 데이터 사이의 거리를 계산한 후 평균을 구하여 K개의 지점을 나눈다. 세 번째로 나뉜 각각의 K구역으로부터 데이터의 평균값을 구해 새로운 중심점으로 지정한다. 이후 두 번째와 세 번째의 과정을 전체 데이터가 가진 분산값이 최소화될 때까지 반복한다. 아래의 K-means 클러스터링 알고리즘 식은 (1)과 같다.

$$\operatorname{argmin}_s \sum_{i=1}^k \sum_{X \in S_i} \|X - \mu_i\|^2 \quad (1)$$

Where, X denotes the given data set, μ_i denotes the mean of points in S_i , S_i denotes the i-th cluster, $\|\cdot\|$ denotes a Euclidean distance and S denotes a set of S_i

3. 연구결과

본 연구에서는 2011년부터 2020년 간 네이버 뉴스에 등록된 간호장교 관련 기사의 댓글을 파이썬 크롤러를 사용해 수집하였다. 뉴스기사의 글 특성상 작성 언론사 및 기자의 영향을 많이 받기 때문에 네이버 뉴스 매체에

선정된 일간지 기사를 선택했다. ‘간호장교’, ‘간호사관’이라는 키워드를 포함되어있는 모든 기사를 카테고리에 무관하게 선정 후 중복된 글을 제거하였다. 크롤링한 총 기사 데이터의 수는 11,433건이다.

수집한 데이터는 비정형 텍스트 데이터의 전처리 과정을 통해 Doc2Vec의 입력으로 사용될 수 있는 단어 토큰의 리스트 형태로 가공하였다. 토큰의 리스트 형태란 각 문서를 형태소 분석을 통해 문장의 각 요소를 나누어 토큰화한 형태를 말한다. 또한 글 전체의 토큰을 Doc2Vec의 사전에 등록해 학습 가능하도록 한다. 기사 데이터 내에 포함된 의미를 갖지 않는 문구 혹은 특정 기사 사이트의 이름은 불용어로 처리하여 전체 데이터 내에서 제거하였다. 그후 파이썬 라이브러리 KoNLPy 패키지의 형태소 분석기를 사용하여 토큰화시켰다[18]. 토큰화 되어진 데이터는 파이썬 라이브러리 Gensim의 Doc2Vec의 입력 데이터로 사용하였다[19].

3.1 문서기반 벡터와 K-means 클러스터링 분석

전처리된 비정형 텍스트 데이터를 분석이 가능한 벡터로 정형화된 데이터로 변환하는 작업을 수행하기 위해 Doc2Vec을 구현하였다. 간호장교 뉴스기사 데이터 특성상 각 문서간의 데이터 길이가 다르며, 여러 의견이 다양하게 나뉜다는 점에서 DM모델을 사용하여 학습하였으며, 평균 문서길이에 따라 Window size를 10, Train

epoch은 50, 차원의 수는 100으로 그리드 서치(Grid search) 기반으로 최적화된 모수를 선택하였고, DM모델의 구현 결과 각각의 문서는 100차원의 벡터로 재현되었다. Fig. 3에서 보는 바와 같이 K-means 클러스터링 알고리즘의 반복을 통해 데이터를 K개의 군집으로 나누어 간호장교에 관한 인식을 군집화하기 위해 가장 적정의 군집개수인 K를 선택하는 엘보우기법(elbow method)[20]의 데이터 최적화 과정을 보여준다.

엘보우기법이란 클러스터 수의 증감에 따르는 데이터에 대한 설명력 변화, 클러스터의 오차를 비교하여 정량적으로 보여주는 지표이다. 따라서 클러스터를 추가함으로써 데이터에 대한 설명력이 어느 정도 증가하는지, 클러스터간의 오차제곱합의 크기가 어느 정도 줄어드는지 확인하여 데이터에 적합한 클러스터의 수를 채택함으로써 연구를 진행하였다. 4개의 군집을 선택했을 시 오차제곱합의 크기가 줄어드는 비율이 상대적으로 크기 때문에 전체 클러스터를 4개로 선정하였다. K-means 클러스터로 군집화한 후 100차원의 데이터를 2차원으로 차원을 축소하여 Fig. 4에서 보는 바와 같이 시각화하였으며 Table 1은 전체 데이터를 군집별로 분류한 데이터의 수를 보여준다.

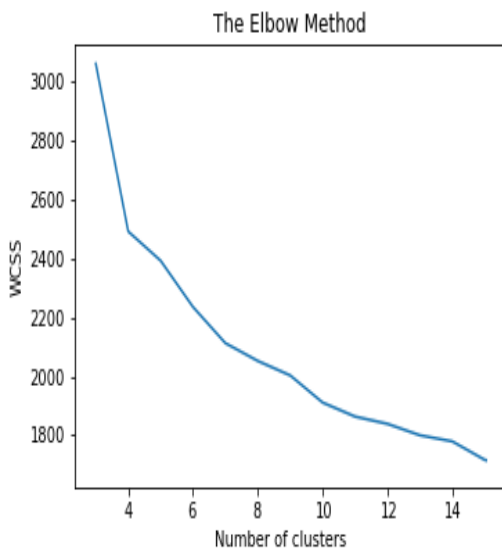


Fig. 3. Elbow method

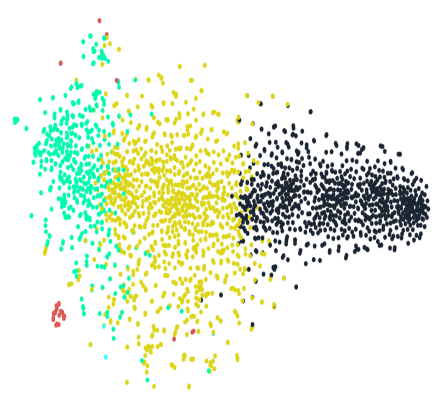


Fig. 4. Cluster visualization

Table 1. Elbow method

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
n	845	491	1058	35

3.2 동시발생행렬 기반의 단어 식별

분류되어진 군집별로 간호장교에 관한 어떠한 인식들이 나누어지는지 확인하고자, 단어간의 동시발생행렬을

기반으로하는 워드네트워크를 설계하였다. 군집별 문서에서 명사만을 추출하여 토큰화시킨 각각의 단어별 동시에 발생한 빈도를 통해 행렬인 단일 단어기반 동시 발생행렬을 만들었으며, 단어간의 빈도수는 단어가 가지는 가중치(Weight) 역할을 수행한다. 이 행렬 중 ‘간호장교’를 핵심 단어로 설정한 후 1차로 인접한 키워드를 선정하 뒤, 각 키워드에 해당하는 인접 키워드를 반복 추출하는 형식으로 설계하였다. 따라서 생성된 워드 네트워크는 간호장교라는 단어 기반의 2차 인접한 단어를 제시하는데 이는 Fig. 5에서 보는 바와 같다. 시각적 편의를 위해 1차로 분류되어진 키워드의 색을 다르게 지정하였다.

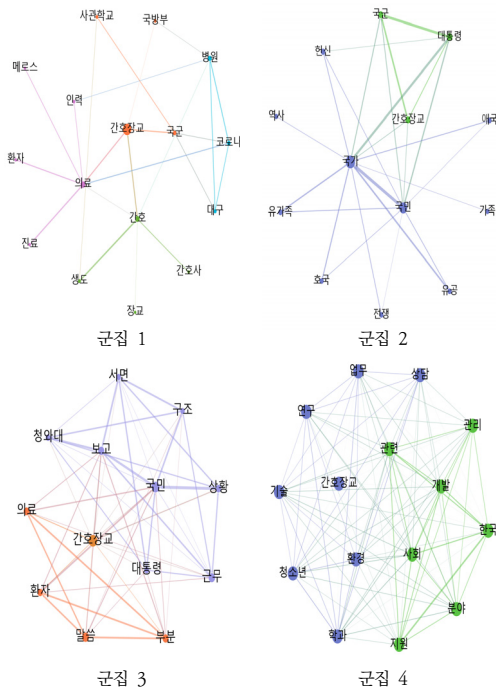


Fig. 5. Word networks based on clustering

3.2.1 군집 1

군집1은 국간사 출신 간호장교의 역할 및 코로나19와 메르스 사태시 간호지원이 연관되어 있는 군집이다. 이는 간호교육 및 군사훈련을 받고 임관한 국간사 출신 간호장교들이 국가적인 의료위기 상황이 발생했을 때 군을 대표하여 인력지원을 실시하고 국민을 위해 환자를 진료한다는 국민들의 인식을 의미한다.

3.2.2 군집 2

군집2는 전쟁 중 전투력을 지원하는데 간호인력이 매

우 중요한 역할을 하였고 그들의 헌신이 군 전투력 향상 및 우리나라의 안보에 큰 영향을 미쳤다는 국민들의 인식을 의미한다. 6.25전쟁 초기에는 여자의용군 중 일부를 간호인력으로 양성하였으나 대통령 지시로 1951년부터 육군병원 부설 간호학교를 창설하여 간호장교를 임관시켰으며, 이는 전투력 보존을 위해 기여하였고, 현 국간사의 모태인 육군간호학교로 이어졌다.

3.2.3 군집 3

군집3은 2014년에 발생한 세월호 사건과 관련되어 있다. 청와대에 근무하며 대통령을 진료하며 건강관리 임무를 수행하는 간호장교가 보고하는 내용의 투명성이 국민에게 요구받을 수 있고 정치적인 입장을 요구받을 수 있는 상황이 발생할 수 있다는 것을 의미한다.

3.2.4 군집 4

군집4는 국간사를 지원하려는 일반학생 및 학부모의 입장과 연관된 군집이다. 청소년 직업상담간 간호장교라는 직업이 전문기술을 가진 직업으로 분류될 수 있는데 대학진로 및 졸업 후 사회에서 활용할 수 있는 분야 등에 대한 의견을 포함하고 있다.

4. 결론

본 연구는 국간사와 간호장교에 대해 국민들이 온라인 상에서 10여년 동안 개진한 방대한 양의 데이터를 크롤링하여 텍스트마이닝 기법으로 군집화하여 분석했으며 그 의미는 크게 두 가지로 볼 수 있다. 첫째, 본 연구는 대국민 신뢰도 증진은 무형 전투력을 향상시킨다는[21] 이론을 기초로 하여, 기존의 군대 내부 인원들에 대한 제한된 설문조사를 통한 군병원 진료 만족도 정도만 분석하는 틀을 뛰어넘어 군 외부에서 국간사와 간호장교를 바라보는 인식을 가시화하여 분석한 의미를 가진다.

둘째, 수집된 데이터는 빅데이터라고 불릴 수 있는 많은 양의 비정형데이터이며 이를 분석하여 형성된 4개의 군집은 각각이 빈도수가 높을 뿐만 아니라 상호 연관성을 고려하여 분석되었기 때문에 신뢰성이 보장된 군집 의견으로 전체적으로 볼 때 군 외부 인식이 긍정적이라는 것이다. 우리는 국간사가 평시 전염병과 같은 국가위기 상황 개선과 전시 군 전투력 보존에 이바지하는 정예장교를 양성하는 사관학교로서의 역할을 수행하고 있다는 인식이 형성되어 있다고 볼 수 있다.

추가하여 대학 진학과 직업이라는 관점에서 청소년과 학부모들이 국간사와 간호장교를 바라보는 인식이 하나의 큰 군집으로 형성되었다는 것을 확인한 것도 의미있는 결과라고 볼 수 있으며 이러한 군집을 통한 분석 결과는 현상 진단과 개선방향 설정에 활용될 수도 있을 것이다.

한편 본 연구에서 사용한 방법 외에도 비정형 데이터 분석방법과 유사하거나 차별화된 군집화와 차원축소 분석기법 등이 있다. 이러한 방법론의 시도를 통해 새로운 분석결과를 얻을 수 있는 가능성은 열려 있으며, 보다 구체화된 의미론적 분석 수행이 가능할 수 있다. 본 연구를 시발점으로 육·해·공군, 해병대를 포함한 우리 군을 바라보는 대국민 인식도와 흐름을 진단하는 다양한 연구들이 이어지길 기대한다.

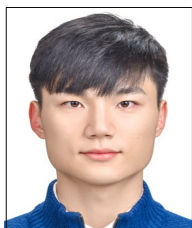
References

- [1] N. Hwang, "Armed forces daegu hospital covid-19 experience", *Journal of Military Nursing Research*, Vol.38, No.2, pp.68-73, 2020.
DOI: <https://doi.org/10.31148/kimnr.2020.38.2.68>
- [2] C. Olston, M. Najork, "Web crawling", *Foundations and Trends in Information Retrieval*, Vol.4, No.3, pp.175-246, 2010.
DOI: <https://doi.org/10.1561/15000000017>
- [3] A. Kao, S. R. Potee, "Natural language processing and text mining", *Natural language processing and text mining*, Springer Science & Business Media, 2007.
- [4] C. Li, J. Bai, L. Zhang, H. Tang, Y. Luo Y, "Opinion community detection and opinion leader detection based on text information and network topology in cloud environment", *Information Sciences*, Vol. 504, pp.61-83, 2019.
DOI: <https://doi.org/10.1016/j.ins.2019.06.060>
- [5] S. Rill, D. Reinel, J. Scheidt, R. V. Zicari, "Politwi: Early detection of emerging political topics on twitter and the impact on concept-level sentiment analysis", *Knowledge-Based Systems*, Vol. 69, pp. 24-33, 2014.
DOI: <https://doi.org/10.1016/j.knosys.2014.05.008>
- [6] Zhang, H. Wang, L. Cao, W. Wang, F. Xu, "A hybrid term-term relations analysis approach for topic detection", *Knowledge-Based Systems*, Vol. 93, pp.109-120, 2016.
DOI: <https://doi.org/10.1016/j.knosys.2015.11.006>
- [7] D. Yan, E. Hua, B. Hu, "An improved single-pass algorithm for chinese microblog topic detection and tracking", *IEEE International Congress on Big Data*, pp.251-258, 2016.
DOI: <https://doi.org/10.1109/BigDataCongress.2016.39>
- [8] H. Lin, B. Sun, J. Wu, H. Xiong, "Topic detection from short text: a term-based consensus clustering method", *13th International Conference on Service Systems and Service Management*, pp.1-6, 2016.
DOI: <https://doi.org/10.1109/ICSSSM.2016.7538624>
- [9] Q. Le, T. Mikolov, "Distributed representations of sentences and documents", *International conference on machine learning*, pp.1188-1196, 2014.
DOI: <https://doi.org/10.48550/arXiv.1405.4053>
- [10] T. Staykovski, A. Barrón-Cedeno, G. San Martino, P. Nakov, "Dense vs. sparse representations for news stream clustering", *Text2Story ECIR*, pp. 47-52, 2019. Available From: <https://ceur-ws.org/Vol-2342/paper6.pdf>
- [11] I. M. Rădulescu, C. O. Truică, E. S.Apostol, C. Dobre, "Enhancing scientific collaborations using community detection and document clustering", *IEEE 16th International Conference on Intelligent Computer Communication and Processing*, pp.43-50, 2020.
DOI: <https://doi.org/10.1109/ICCP51029.2020.9266267>
- [12] J. J. Chu, H. C. Chung, O. S. Hyeon, Y. J. Hyeon, Y. S. Park, "A study on nursing officer's nursing image", *Journal of Military Nursing Research*, Vol 16, pp.5-29, 1998.
- [13] J. S. Na, J. Y. Lee, "A study on relationship between patient perceived nursing officer's communication style and patient reliability, satisfaction", *Journal of Military Nursing research*, Vol. 34, No. 1, pp.94-102, 2016.
DOI: <https://doi.org/10.31148/kimnr.2016.34.1.94>
- [14] C. H. Woo, J. Y. Park, H. Kim, "Effect of calling, goal orientation, collective efficacy on organizational socialization among korea armed forces nursing academy cadets", *Journal of Military Nursing Research*, Vol. 38, No. 1, pp.67-78, 2020.
DOI: <https://doi.org/10.31148/kimnr.2020.38.1.67>
- [15] B. Lee, "Experience of covid-19 patients care in infectious diseases specialized hospital in daegu", *Journal of Military Nursing Research*, Vol. 38, No. 2, pp.74-79, 2020.
DOI: <https://doi.org/10.31148/kimnr.2020.38.2.74>
- [16] T. Mikolov, K. Chen, G. Corrado, J. Dean, "Efficient estimation of word representations in vector space". *International Conference on Learning Representations*, Jan. 2013.
DOI: <https://doi.org/10.48550/arXiv.1301.3781>
- [17] J. A., Hartigan, M. A. Wong, "Algorithm as 136: A k-means clustering algorithm", *Journal of the Royal Statistical Society. Series C(Applied Statistics)*, Vol 78, No. 1, pp.100-108, 1979. Available From: <https://www.jstor.org/stable/i316073>
- [18] E. L. Park, S. Cho, "KoNLPy: korean natural language processing in python", *Annual Conference on Human and Language Technology*, pp.133-136, 2014. Available From: <https://koreascience.kr/article/CFKO201408355727285.pdf>

- [19] R. Řehůřek, P. Sojka, "Gensim—statistical semantics in python", *Retrieved from Gensim.org*. 2011.
<https://www.fi.muni.cz/usr/sojka/posters/rehurek-sojka-scipy2011.pdf>
- [20] T. M. Kodinariya, P. R., Makwana, "Review on determining number of Cluster in k-means clustering", *International Journal of Advance Research in Computer Science and Management Studies*, Vol. 1, No 6, pp.90-95, 2013. Available From:
<https://www.researchgate.net/publication/313554124>
- [21] D. G. Cha, "Study on the strategic culture and behavior of marine corps", *Journal of Korean-Japanese Military and Culture*, Vol. 29, pp. 85-113, 2020.
DOI: <https://doi.org/10.47563/KJMC.29.4>
-

이 한 인(Hanin Lee)

[정회원]



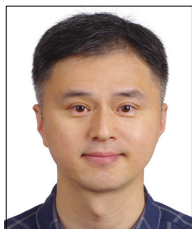
- 2021년 3월 : 육군사관학교 수학과 (이학사)
- 2021년 6월 ~ 2022년 6월 : 21보병사단 소대장
- 2022년 6월 ~ 현재 : 21보병사단 작전장교

<관심분야>

머신러닝, 텍스트마이닝, 통계

차 진 호(Jinho Cha)

[정회원]



- 2007년 8월 : 미국 텍사스A&M대학원 산업공학과 (산업공학석사)
- 2015년 8월 : 미국 클렘슨대학원 산업공학과 (산업공학박사)
- 2017년 7월 ~ 2021년 8월 : 육군사관학교 수학과 조교수
- 2021년 9월 ~ 현재 : 국방기술진흥연구소 연구원

<관심분야>

통계, 최적화, 텍스트마이닝