

COVID-19 팬데믹 상황에 대한 국내 트위터 데이터 분석

용혜련¹ · 황현석^{2*}¹선임연구원, 효성 ITX^{2*}한림대학교 경영학과 교수

Mining Twitter Texts for the COVID-19 Pandemic in Korea

Hye-Ryeon Yong¹ · Hyun-Seok Hwang^{2*}¹Senior Researcher, Hyo-Sung ITX, Seoul, Korea^{2*}Professor, Department of Business Administration, Hallym University, Chun-Cheon 24252, Korea

[요약]

COVID-19 팬데믹은 정신 건강에 상당한 영향을 미쳐 스트레스, 불안, 우울증, 고립감의 수준을 증가시켰다. 이러한 정신 건강 문제를 극복하는 것은 개인의 신체적, 정서적 행복에 미치는 부정적인 결과를 예방하기 위해 매우 중요하다. 또한, 정신건강의 문제는 신체적 건강 문제를 효과적으로 관리할 수 있는 능력 저하로 이어질 수 있기 때문에 전반적인 공중보건을 위해 정신건강 문제를 해결하는 것이 중요하다. 본 연구에서는 소셜 미디어에서 나타나는 대중의 심리상태가 어떠한지 살펴보기 위해 트위터에 게시된 트윗을 수집하여 소셜 미디어에서 발생하는 커뮤니케이션의 구조 및 특성을 파악하고자 하였다. 이를 위해 트위터 데이터를 수집하는 코드를 작성하고 수집된 텍스트의 분석을 위해 첫째, 새로운 다차원의 감성사전을 구축하고 둘째, 코로나 상황에 따라 대중의 심리적 반응은 어떻게 변화하는지 분석하며 셋째, 대유행 기간별, 감정별 시계열 추이를 어떠한지 트위터 데이터를 통해 실증적으로 검증하고자 하였다. 분석결과 코로나 상황의 변화가 트위터에서 감정을 표출하는 데 있어 대중에게 영향을 미치고 있음을 확인할 수 있었다. 이는 연구결과 소셜미디어에 나타난 텍스트의 분석을 통해 방역 대응 정책의 지표로써 활용될 수 있는 가능성을 보여준다.

[Abstract]

The COVID-19 pandemic has had a significant impact on people's mental health, increasing their levels of stress, anxiety, depression, and isolation. Overcoming these mental health issues is crucial to prevent negative consequences on an individual's physical and emotional well-being. Additionally, addressing mental health issues is important for maintaining overall public health, as they can reduce one's ability to effectively manage physical health problems. In this study, we sought to understand the structure and nature of social media communication by collecting tweets posted on Twitter to explore the public's mental state. Accordingly, we wrote a code to collect Twitter data and analyze the collected texts to 1) build a new multidimensional emotional dictionary, 2) analyze how the public's psychological reactions change according to the COVID-19 situation, and 3) empirically verify the time series trend of emotions during the pandemic through Twitter data. The results showed that the analysis of texts on social media can be used as an indicator for quarantine response policies.

색인어 : 코로나19, 정신건강, 트위터, 빅데이터, 감성분석**Keyword** : COVID-19, Mental Health, Twitter, Big Data, Sentiment Analysis<http://dx.doi.org/10.9728/dcs.2023.24.5.1111>

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 25 February 2023; Revised 16 March 2023

Accepted 03 April 2023

***Corresponding Author; Hyun-Seok Hwang**

Tel: +82-33-248-1835

E-mail: hshwang@hallym.ac.kr

I. 서론

코로나19 팬데믹으로 인해 전 세계적으로 심각한 사회적 피해가 발생시키고 있다. 전 세계적으로 수백만 명이 사망하고 더 많은 사람들이 장기적인 건강 합병증을 앓고 있다. 경제적으로는 여행, 숙박업, 소매업과 같은 산업에서 일자리 손실, 사업장 폐쇄, 경기 침체를 초래하였고 교육 측면에서 전 세계 교육 시스템이 혼란에 빠지면서 휴교령과 원격 수업이 이어졌다. 이로 인해 기존의 교육 불평등이 확대되고 학생과 교육자에게 새로운 도전 과제가 생기게 되었다.

코로나 장기화로 인한 대중의 피로도가 누적되면서 이로 인한 우울, 분노, 무기력함 등의 감정이 고조되며 코로나 블루(corona blue), 코로나 레드(corona red), 코로나 블랙(corona black)과 같은 신조어가 등장하였다. 가장 먼저 생성된 코로나 블루란 ‘코로나’와 ‘우울감(blue)’이 합쳐진 신조어로, 코로나의 확산으로 일상에 큰 변화를 겪으며 생긴 우울감이나 불안감을 뜻한다[1]. 코로나 레드 는 우울을 넘어서 분노를 느끼는 상태이며, 코로나 블랙은 장기화되는 코로나로 인한 절망감·아담·무력감 등을 느끼는 심리적 상태로 코로나 블루와 코로나 레드에서 파생되었다[2]. 코로나 블루는 감염에 대한 불안, 사회적 거리두기와 집에 머물기로 인한 타인과의 관계 단절, 경기 침체로 인한 소득 감소, 고용 위기로 인한 일자리 감소, 어린 자녀들의 돌봄 및 교육 부담 등과 같은 다양한 이유에서 발생하고 있다[3]-[5].

소셜 빅데이터 분석은 정부가 대중의 정서와 의견에 대한 인사이트를 얻을 수 있는 다양한 기회를 제공한다. 대중의 관점을 이해함으로써 정부는 보다 효과적인 커뮤니케이션 전략을 수립하고, 대중의 요구에 맞게 정책을 조정하며, 전반적인 평판을 개선할 수 있다.

본 연구에서는 국내 COVID-19 팬데믹 상황에 따라 소셜 미디어에서 나타나는 공중의 심리상태는 어떠한지 살펴보고자 하였다. 소셜 미디어에서 발생하는 커뮤니케이션의 구조 및 특성을 파악하고자 하였으며 다음의 3가지 관점에서 분석 데이터를 정하였다. 첫째, 위기 상황에서 공중의 커뮤니케이션을 내용을 파악하고 이에 따른 적절한 대응 방안을 모색할 수 있는 데이터가 필요하다. 둘째, 팬데믹 상황에서의 공중의 정신건강과 심리를 파악하기 위해 정제되지 않은 빅데이터를 활용해 볼 필요가 있다. 셋째, 감염병 위험에 대한 불안감이 급증한 상황에서 실시간으로 업로드되는 소셜 미디어를 활용해 볼 필요가 있다. 이를 위해 본 연구에서는 소셜 미디어 플랫폼인 트위터(twitter)를 활용하고자 한다.

2장에서는 소셜미디어를 활용한 재난 및 위기 관리 관련 연구와 텍스트 마이닝을 이용한 헬스케어 연구에 대한 선행 연구를 살펴보고 3장에서는 본 연구의 연구설계를 위한 연구 문제, 연구 프레임워크를 제시하였다. 3장의 연구를 실증분석하기 위해 4장에서는 수집된 데이터의 분석과 결과를 제시하였으며, 5장에서 본 연구의 의의와 한계점을 제시하였다.

II. 소셜 빅데이터를 활용한 재난, 위기, 헬스케어 연구

2-1 소셜 미디어를 활용한 재난 위기관리

최근 긴급 상황과 위기 대응에서 소셜 미디어가 커뮤니케이션 수단으로써 결정적인 역할을 하는 요소로 진화하고 있다[6],[7]. 소셜 미디어는 조직체나 기관에서 부각되지 않은 실수나 문제점을 쉽게 쟁점화하거나 사회적 이슈에 대한 정치참여를 촉진시키는 도구로 그 영향력을 행사하고 있다[8]. 나아가 마이크로 블로깅 서비스인 트위터로 강화된 즉시성과 전파성은 기존 뉴스 유통 방식의 변화까지 주도하고 있는 실정이다[9].

논문[10]의 연구에서는 코로나로 인해 이로 떠오른 마스크 5부제에 대한 온라인 뉴스 기사와 카페 게시글을 분석하여 대중의 의제를 파악하고, 정부의제로 진화하는 정책의제 설정 과정에서 참고로 활용될 수 있음을 제시하였다. 논문[11]은 재난이 발생했을 경우 소셜 미디어를 통해 피해를 감지할 수 있음을 주장하였다. 소셜 미디어에서 발생한 메시지의 지리적 위치를 추출하여 심각한 피해 지역과 가벼운 피해 지역을 구분하고 구조 활동의 우선순위를 정할 수 있음을 제시하였다. 논문[12]의 연구에서는 2017년에 발생한 강릉·삼척 산불 사태에 나타난 위험 커뮤니케이션의 네트워크 및 구조 패턴을 분석하였으며 정보이동과 관계구조 파악을 통해 커뮤니케이션의 구조와 행위자 간 상호작용을 살펴보았다. 논문[13]의 연구는 국가별로 코로나19 관련 위기 커뮤니케이션을 위해 소셜미디어를 어떻게 활용하는지를 분석하고, 게시글의 유형(감염병 정보, 행동지침, 심리적 소통)에 따른 이용자 반응(Engagement) 수준을 비교 분석하였다.

2-2 헬스케어와 텍스트 마이닝

텍스트 마이닝은 텍스트 기반 데이터로부터 자연어 처리(Natural Language Preprocessing)를 통한 유용한 지식 발견기법을 말한다[14]. 논문[15]의 연구에서는 한국학술지인용색인(KCI)에서 ‘감정노동’이라는 주제가 포함된 1,465건의 검색된 논문을 심층적으로 분석하여 감정노동 연구 동향을 파악하였다. 논문[16]은 Covid-19 팬데믹이 항공산업에 미친 영향을 살펴보고자 국내 뉴스 기사 데이터를 활용하여 키워드 트렌드 분석을 진행하였다.

한편 텍스트 마이닝을 헬스케어 분야에 접목한 다수의 연구가 진행되었는데, 논문[17]은 소셜 미디어에서 ‘미세먼지’ 키워드로 검색한 결과를 수집한 후, 매스미디어와 소셜미디어의 언급을 비교하였다. 분석결과, 건강 행위와 관련한 정보가 소셜 미디어에서 월등한 수준으로 높게 언급되었음을 밝혔다. 논문[18]의 연구에서는 중국 소셜 미디어인 Weibo를 대상으로 중국의 가족계획 정책에 대한 커뮤니케이션과 주요 이슈는 무엇인지 파악하고자 하였으며 이를 위해 Latent

Dirichlet Allocation 분석을 실시하였다. 논문[19]는 전 세계적 이슈인 시리아 난민 위기와 이민에 대한 대중의 의견과 감정을 파악하기 위해 터키어와 영어로 된 약 2백만여 개의 관련 트윗을 분석하였다.

III. 연구 문제와 연구 설계

3-1 소셜 빅데이터를 활용한 감성분석 연구문제

본 연구에서는 앞의 이론적 고찰을 토대로 트위터 데이터를 활용하여 장기화된 COVID-19 팬데믹에 대한 대중들의 심리적 반응을 살펴보고, 대유행 상황에서 논의되는 주요 의제는 무엇인지 파악해보고자 한다. 본 논문에서 설정한 연구 문제는 다음과 같다.

<연구문제 1> COVID-19와 관련된 국내 트위터에서, 상황의 변화에 따라 대중의 심리적 반응은 어떻게 바뀌는가?

<연구문제 2> COVID-19와 관련된 국내 트위터에서, 부정적 감정 중 슬픔과 분노를 표현하는 데 있어 대유행 기간별 어떠한 시계열 추이를 보이는가?

본 연구에서는 트윗에 사용된 텍스트 데이터를 연구 대상으로 하였으며, 특정 키워드(“#코로나”, “코로나”, #COVID)

를 포함하는 트윗을 2020년 11월부터 2021년 10월 동안 스크래핑(scraping)하였고, Twitter API와 R 프로그램을 활용하였다. 아래의 그림 1은 1년간 수집된 트위터 데이터로 x축은 수집시기이며 y축은 수집된 트윗 수를 나타낸다.

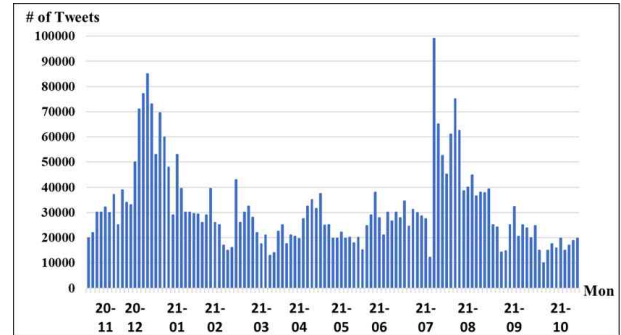


그림 1. 코로나19와 관련된 트윗의 빈도
Fig. 1. Frequency of Tweets related to Covid-19

감성분석 대상이 되는 트위터 데이터는 총 3백 7십만 건이며, 이 중 1백만 건을 사용하였다. 텍스트 데이터가 완전히 중복되는 트윗은 업로드 시간이 달라도 분석에서 제외하였다. 제외한 이유는 다음과 같이 두 가지로 논할 수 있다.

첫째, 스팸(spam)활동에 대한 고려이다. 스팸 게시물은 사람들이 관심 없거나 원하지 않는 메시지인 경우가 대부분이다. 이러한 스팸 트윗은 대중의 심리적 변화를 파악하는데 저

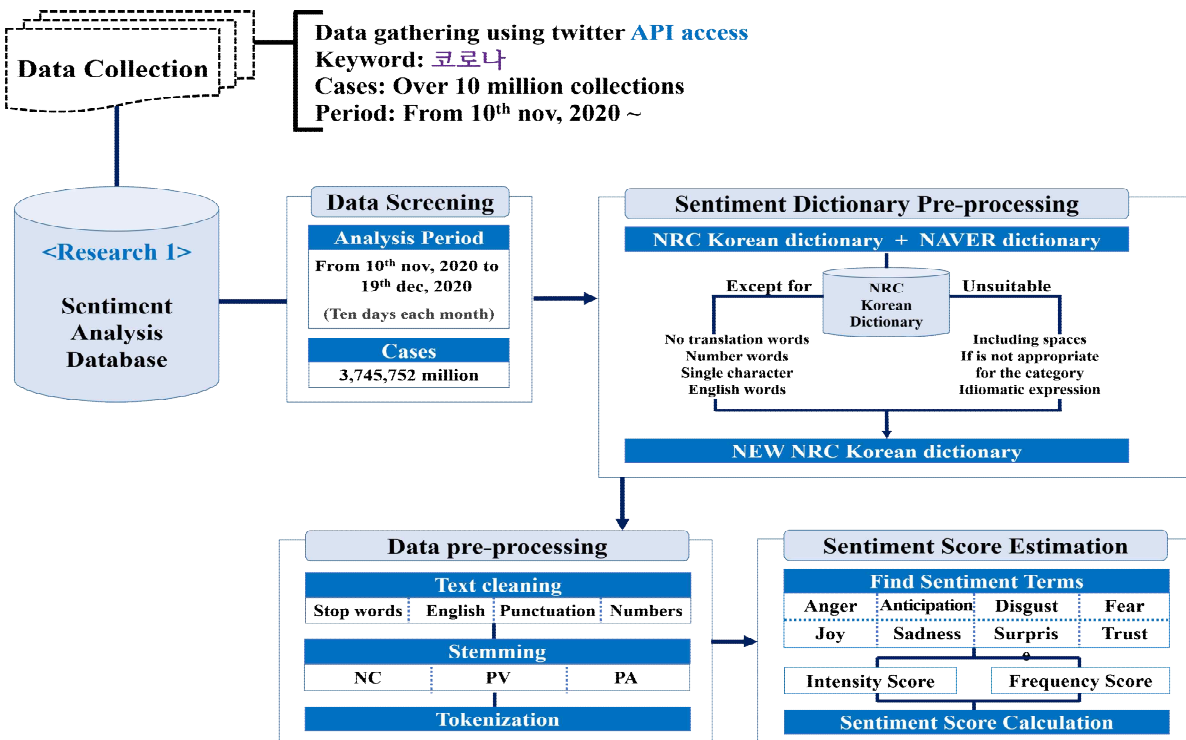


그림 2. 연구 프레임워크
Fig. 2. Research framework

해 요인이 될 수 있다.

둘째, 리트윗 영향력을 최소화하고자 하였다. 트위터는 세계 각국에서 사용하는 최대 소셜 미디어 플랫폼으로, 일반 대중뿐만 아니라 정치인, 연예인, 인플루언서 등 각계각층의 영향력 있는 유명인사도 다수 사용하고 있다[20],[21]. 따라서 리트윗된 데이터를 모두 반영하는 경우 특정 단어가 지나치게 많이 포집되는 문제가 발생할 수 있다.

3-2 소셜 빅데이터를 활용한 감성분석 연구 프레임워크

앞서 논의된 연구자료를 토대로 어휘 기반의 감성분석을 실시하였다. 본 논문의 연구절차는 그림 2와 같다.

위에 제시된 바와 같이, 연구결과 생성을 위해 감성사전의 이용이 필수적이다. 본 연구의 감성분석에 사용된 사전은 NRC(NRC word-emotion association lexicon) 사전을 활용하였다. NRC 사전은 14,182개의 영어 단어를 중심으로 한국어 포함 약 100여 개국 언어에 대한 감성 단어를 제공한다. 본 연구에서는 NRC 한국어 버전 사전과 네이버 사전의 결합을 통해 어휘기반 감성분석의 핵심이 되는 감성사전을 재구축하였다.

NRC 사전은 기존의 긍정-부정 차원을 넘어서, ‘분노(anger)’, ‘기대(anticipation)’, ‘혐오(disgust)’, ‘공포(fear)’, ‘기쁨(joy)’, ‘슬픔(sadness)’, ‘놀람(surprise)’, ‘믿음(trust)’ 총 8가지의 감성 차원과 유사도 점수(intensity score)를 제공한다. NRC 사전은 각 차원에 대한 단어들의 차별화 된 유사도 점수를 가지고 있어 단어가 어느 차원에서 얼마나 강하게 속하는지를 알 수 있다(최건우 외, 2019). 유사도 점수는 0~1 사이의 값을 가지며, 1에 가까울수록 속한 차원과 강한 연결 관계이다. 예를 들어 ‘우울’이라는 단어는 NRC 사전에 따라 부정적이면서 ‘분노’와 ‘슬픔’을 가지고 있는 단어로 분류할 수 있으며, 슬픔 0.621점, 공포 0.312점과 같이 유사도 점수를 계산할 수 있다. 또한 ‘우울’, ‘우울증’, ‘우울한’과 같이 다양한 형태로 표현되어 있다. 본 연구에서는 차원에 대한 연결 강도를 알 수 있는 유사도 점수에 단어의 표현 여부만 알 수 있는 빈도 점수를 추가하여 총 두 가지의 감성사전 기반 점수를 구축하였다.

체계적인 감성분석을 위해 NRC 한국어 버전 사전에서 제공되는 8차원의 약 1만여 개 감성 단어의 영어 단어와 한국어 해석이 알맞게 매치되었는지 확인하고자 하였다. 이를 위해 R 프로그램을 활용하여 네이버 사전에서 NRC 사전에 등재되어 있는 영어 단어를 한국어로 번역하고 이를 자동으로 수집하였으며, 과정을 통해 기존에 NRC 사전에서 번역되지 않거나 잘못 번역된 한글 단어를 새롭게 교체하였다. 네이버 사전으로 한 번 더 보완했음에도 불구하고 한글로 번역되지 않은 단어 477개와 띄어쓰기가 포함된 단어 2,023개, 한 글자 단어 241개, 영어 표현 단어 26개, 숫자가 포함된 단어 2개는 제외하였다. 표 1에서는 제거된 단어의 일부를 예시로 보여주고 있다.

표 1. NRC 사전에서 제외된 단어들

Table 1. Words excluded from NRC

Remove	Examples
Inappropriate emotional words	Anticipation-anger
Political words	Trust-president
Religious words	Joy-Jesus
Words related to unique roles	Disgust-man
Spam-related words	Trust-sex
Onomatopoeic/ mimetic words	Anger-Ah Ah
Words related to time	Joy-time
Animal-related words	Disgust-donkey
Words related to a country	Joy-Japan
Emotionally incorrect words	Fear-chicken

이러한 전처리 과정을 거쳐 감성분석을 수행하기 위한 새로운 감성사전을 구축하였다. 기존 NRC 사전과 새롭게 구축된 사전의 구성은 표 2와 같다.

표 2. 새로운 감정어 사전

Table 2. New emotional dictionary

Dimension	Cases	English Words	Pre-processed Korean Words
Anger		1,250	1,039 (15%)
Anticipation		842	584 (8%)
Disgust		1,060	824 (12%)
Fear		1,483	1,300 (19%)
Joy		691	795 (11%)
Sadness		1,195	990 (14%)
Surprise		535	388 (6%)
Trust		1,234	1,027 (15%)
Negative		3,338	-
Positive		2,317	-
Total		13,945	6,947 (100%)

이후 감성분석에 사용될 트위터 텍스트 데이터를 단어 수준의 분석으로 수행하기 위해 자연어 처리를 진행하였다. 먼저 문장 단위의 트위터 텍스트를 단어 단위로 쪼개기 위해 형태소 분석을 실시해야 하는데, 이러한 형태소 분석을 하고자 할 때, 기준이 될 사전이 필요하다. 분석할 문장에 포함된 단어들 중 특정 사전에 알맞은 형태(품사)로 포함되어 있어야만 원하는 형태의 단어를 추출할 수 있기 때문이다. 형태소 분석을 위해 R의 KoNLP(Korean natural language process) 패키지와 NIADic을 활용하였다.

IV. 실증분석 및 결과

4-1 실증분석 과정

사전기반 감성분석의 단어 단위 분석에서는 먼저 감성분석의 대상이 되는 자료(target)에서 단어(word)를 추출하고, 추출된 단어가 감성 사전(dictionary)에 포함되어 있는지 확인한다. 이후 감성 사전에서 미리 정의된 감정 점수를 바탕으로 단어 및 문장별 감성을 계산하고, 그 결과값으로 분석 대상의 감성을 분류한다. 그림 3은 사전기반 감성분석의 알고리즘을 제시한 것이다.

2020년 11월부터 2021년 10월까지 지난 1년간 트위터에서 수집된 텍스트 데이터를 위와 같은 과정을 거쳐 토큰화하고, 토큰화한 약 1천만 개의 데이터를 앞서 설명했던 새롭게 구축한 감성 사전과 매치시켰다. 토큰 단어와 사전 단어의 매치를 통해 두 가지 감성 점수인 유사도 점수와 빈도 점수를 도출하였고, 이를 활용하여 텍스트 데이터를 일별·월별·시간별 8차원의 감성 점수로 수치화하였다.

```

for w ∈ t do
  while w ∈ d do
    sentiment = ∑(score(wi))
  end while
end for
    
```

w: word, t: target, d: dictionary

그림 3. 어휘기반 분류 알고리즘
 Fig. 3. Lexicon based classification algorithm

사전기반 감성분석의 단어 단위 분석에서는 감성분석의 대상이 되는 자료 *t*에서 단어 *w*를 추출하고, 추출된 단어가 감성 사전 *d*에 포함되어 있는지 확인한다. 이후 감성 사전에서 미리 정의된 감정 점수를 바탕으로 단어 및 문장별 감성 *sentiment*를 계산하여 분석 대상의 감성을 분류한다.

NRC 사전에 따르면 ‘회피’라는 단어는 공포와 슬픔 두 차원에 모두 해당하는 단어이며 사전에서 단어에 미리 부여한 유사도 점수에 따라 공포 0.483점, 슬픔 0.484점의 감성 점수가 매겨진다. <연구문제 1>에서는 공중의 심리상태를 보다 면밀히 살펴보기 위해 감성 단어와 차원 간 연결 강도를 나타내주는 척도인 유사도 점수를 활용하여 분석하였다.

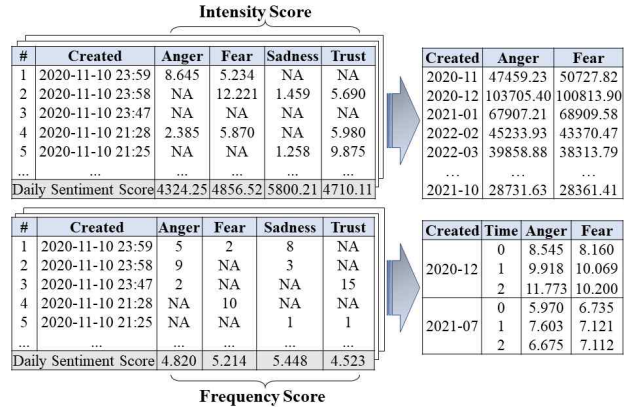


그림 4. 감성분석을 위한 데이터 셋
 Fig. 4. Data set for sentiment analysis

4-2 분석 결과: 연구문제 1

감성분석의 기간은 2020년 11월부터 2020년 10월까지 약 1년으로, 상황에 따른 일별 심리적 변화를 다시 월별로 계산하였다. 월별 감성분석의 순위는 아래의 표 3과 같다.

표 3. 월별 8가지 감정어 사용 순위
 Table 3. Monthly ranking of 8 emotions

	'20/11	'20/12	'21/01	'21/02	'21/03	'21/04	'21/05	'21/06	'21/07	'21/08	'21/09	'21/10
anger	8	1	6	2	2	1	4	5	8	1	7	1
ant	3	8	1	1	8	2	3	6	5	8	8	4
dis	5	2	7	7	3	8	7	2	6	6	1	5
fear	1	6	2	8	7	3	5	1	1	4	4	3
joy	7	5	4	3	4	4	2	3	7	5	6	6
sad	4	7	8	6	5	7	1	8	4	2	5	8
sur	6	3	3	4	6	6	6	7	2	7	2	2
trust	2	4	5	5	1	5	8	4	3	3	3	7

ant: anticipation, dis:disgust, sad: sadness, sur: surprise

아래의 그림 5는 여덟 차원의 일일 감성 유사도 점수와 일일 신규 확진자 수를 나타낸 그림이다. 분석 기간이 길고 감정 차원이 많아 연구결과의 이해를 돕기 위해 누적 그래프로 나타냈다. 그림 5를 살펴보면, 2020년 11월 10일부터 19일 기간에는 ‘공포’ 유사도 점수가 50727.83으로 가장 높으며 ‘믿음’이 49512.44점으로 그 뒤를 따랐다. 반면에 ‘분노’는 4749.5로 가장 낮음을 알 수 있다. 12월은 104640.5로 트위터에서 ‘분노’가 가장 많이 표현되었으며 ‘혐오’ 감성이 103545.2로 그 뒤를 이었다. 가장 낮은 감성 점수는 100606으로 ‘기대’로 나타났다.

2021년 1월에 가장 높은 감성 점수를 보인 단어는 ‘기대’로, 68954.61 점수를 보였으며 ‘공포’가 68909.58점으로 2위의 감성 점수를 보였으며 ‘슬픔’이 66840.31점으로 가장

낮은 점수를 보였다. 2월에는 ‘기대’가 45257.21점, ‘분노’ 차원의 감성은 45233.93으로 높게 표출되었으며 ‘공포’ 감성이 68909.58점으로 가장 낮은 점수를 보였다. 3월에는 ‘믿음’이 40144.81점, ‘분노’가 39858.88점으로 상위에 위치했으며 ‘기대’ 감성이 가장 적게 표현되었다. 4월에는 44113.52점으로 ‘분노’가 가장 많이 표현되었으며 ‘기대’가 43886.9로 두 번째로 높은 감성 차원임을 알 수 있었다. 8위는 ‘혐오’로 43280.55점의 감성 유사도 점수를 보였다. 5월에는 ‘슬픔’이 1등 감성 표현으로 45715.46점이었으며, ‘믿음’이 42847.73점으로 가장 낮은 점수를 보였다.

하반기 시작된 6월과 7월에는 ‘공포’ 감정이 가장 많이 표현되었으며 각각 34746.89점, 73298.08점으로 같은 1위의 감정이지만 월별에 따라 약 두 배의 차이를 보였다. 6월에는 ‘혐오’가 33928.27점으로 ‘공포’ 감정의 뒤를 이었으며, 확진자가 폭발적으로 증가한 7월에는 ‘놀람’이 73177.78점으로 ‘공포’ 감정 다음으로 많이 표현되었다. 8월에는 ‘분노’가 62827.84점으로 가장 높은 감성 표현 빈도를 보였으며 ‘믿음’이 61523.78점으로 2위의 감성 표현 점수가 도출되었다. 9월에는 39245.06점으로 ‘혐오’가 가장 많이 표현되었으며 ‘놀람’과 ‘믿음’이 각각 38551.33점, 38452.05점으로 2, 3위로 나타났다. 마지막으로 10월에는 ‘분노’가 28731.63점으로 가장 높은 감성 표현을 보였으며, ‘슬픔’이 27660.78로 가장 낮은 감성 표현을 보였다.

2020년 11월부터 2021년 10월까지 트위터에서 표현된 월별 감성분석 결과를 종합해보면, 2021년 2월과 3월에 가장 많이 표출된 감정인 ‘기대’와 4월의 ‘믿음’ 감정을 제외하고는 모두 부정적 감정과 관련된 감정들이 그 달에 가장 많이 표출되었음을 알 수 있다. 좀 더 면밀히 살펴보면, ‘분노’는 1위 4번, 2위 2번으로 가장 많이 표현되었다. 다음으로는 ‘공포’가 1위 3번, 2위 1번으로 두 번째로 많이 표출된 감정이었으며 반면에 가장 적게 표현된 감정은 ‘기쁨’으로 5월에 2위로 한 번 표현된 것에 그쳤다.

감성분석 결과에 대한 감성 유사도 점수를 1년 단위로 보면, ‘공포’가 총 625426.9점으로 가장 높은 감성 유사도 점수를 보였다. 이러한 결과는 불확실한 전염병 상황에 따른 감정 표현이라고 볼 수 있다. 그 뒤를 이어 ‘분노’가 624639.3점으로 2위, 3위는 ‘믿음’으로 623704.2점으로 나타났다. 지난 1년간 트위터에서 추출한 코로나 관련 데이터에 대한 감성별 유사도 점수는 Fig. 6과 같다. 이후, 좀 더 면밀한 감성분석을 위해 어떤 감성들이 빈번하게 등장하는지 파악해보고자 하였다. Table 4는 매월 10일에 가장 많이 표현된 상위 다섯 개의 명사, 동사, 형용사 단어의 빈도를 정리한 것이다.

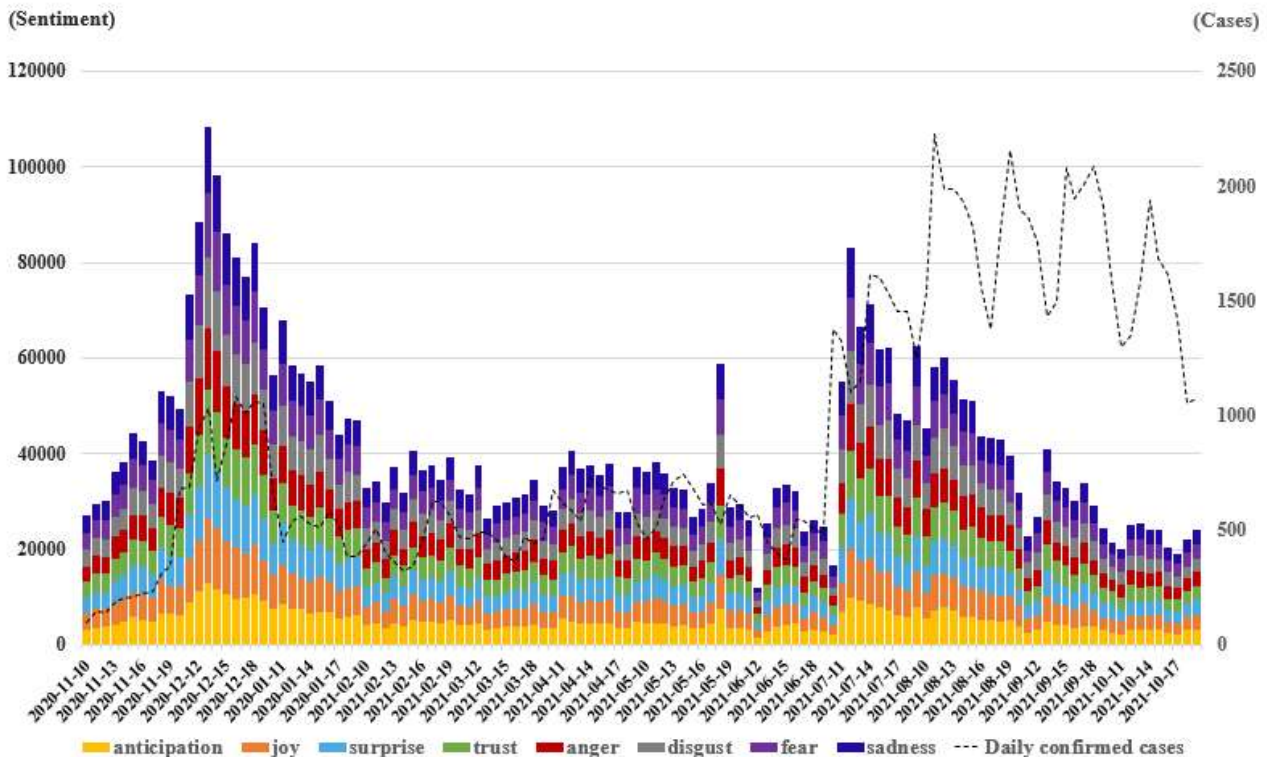


그림 5. 8차원의 감정 유사도 점수와 확진자 수
 Fig. 5. 8D emotional similarity score and the number of confirmed cases

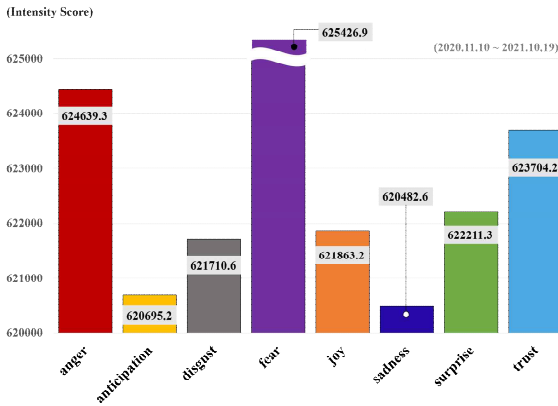


그림 6. 감정별 유사도 점수
Fig. 6. Similarity score by sentiment

표 4. 최상위 5개 명사-동사-형용사
Table 4. Top 5 noun-verb-adjectives

Year / Mon	Noun		Verb		Adjectives	
'20 /11	시작	179	끝나	193	아니	267
	사람	164	힘든	50	안되	53
	상황	148	지나	49	어렵	44
	환자	110	보내	47	무섭	43
	여행	86	미치	39	아프	33
'20 /12	치료제	821	끝나	386	아니	512
	환자	581	걸리	198	안되	112
	감염	539	보내	120	심하	105
	사람	338	조용하	99	무섭	72
	검사	290	지나	94	어렵	72
'21 /01	감염	836	끝나	336	아니	607
	환자	424	기다리	132	어렵	72
	세계적유행병	383	지나	87	안되	59
	치료제	331	조용하	75	다르	57
	의혹	261	따른	72	새롭	44
'21 /02	감염	376	끝나	284	아니	318
	치료제	267	보내	175	어렵	77
	검사	196	걸리	106	즐겁	67
	사람	191	터지	56	다르	48
	접종	156	알리	42	아프	43
'21 /03	치료제	596	끝나	202	아니	239
	감염	304	걸리	105	안되	56
	접종	290	지나	62	아프	53
	환자	278	터지	51	어렵	46
	치료	176	알리	47	심하	36
'21 /04	환자	399	끝나	234	아니	260
	감염	322	걸리	102	심하	48
	치료제	286	지나	73	무섭	47
	사람	243	따르	66	아픈	39
	검사	164	미치	64	빠른	32
'21 /06	치료제	508	끝나	200	아니	264
	감염	379	걸리	99	어렵	86
	접종	360	지나	61	아프	58
	환자	325	보내	57	안되	39
	바이러스	257	따르	46	다른	38
'21 /06	접종	501	끝나	167	아니	197
	감염	349	지나	66	안되	46

'21 /07	환자	343	걸리	65	아픈	40
	치료제	228	다르	45	어렵	31
	사람	176	알리	40	다른	25
	조심	614	걸리	116	아니	161
	감염	223	기다리	112	고맙	86
세계적유행병	건강	201	끝나	104	아프	67
	보내	140	미치	49	심하	43
	환자	138	미치	38	그립	33
	일탈	740	끝나	307	아니	489
'21 /08	환자	643	걸리	253	안되	89
	감염	643	다르	86	아프	81
	검사	358	미치	84	무섭	78
	접종	340	터지	79	심하	77
	환자	351	끝나	178	아니	299
'21 /09	검사	284	걸리	132	외롭	152
	접종	271	지나	62	아프	89
	일탈	237	기다리	57	다르	49
	안녕	203	따르	50	어렵	45
	일탈	258	끝나	178	아니	250
'21 /10	감염	189	걸리	97	아프	54
	사람	189	지나	36	안되	50
	예방	167	기다리	33	무섭	42
	접종	141	미치	30	외롭	36

* The table above is written in Korean to represent the nouns, verbs, and adjectives used in the analysis.

표 4에서 알 수 있듯이, 매월 10일 날짜에 게시된 트위터에서 명사, 동사, 형용사로 토근화 된 상위 5개 단어를 정리하였다. 표의 단어들은 본 연구에서 감성분석을 위해 구축한 사전과 매치된 단어로, 감성 유사도 점수가 계산된 단어이다. 상위 빈출 단어는 감성을 분류하는데 영향을 미칠 수 있으므로 중요한 단어라고 할 수 있다.

4-2 분석 결과 : 연구문제 2

<연구문제 2>에서는 빈도 점수 데이터를 활용하여 슬픔 및 분노와 같은 부정적 감성에 가까운 단어들의 표출이 언제 가장 활발히 나타나는지를 월별로 보았다. 분석 대상은 3차 대유행이 시작된 지난 2020년 12월과 4차 대유행이 도래한 2021년 7월로 선정하였으며 이에 따른 빈도 점수를 계산하였다. 빈도 점수는 한 트윗에 등장한 단어들을 카운트하는 수치로 차원에 대한 단어의 연결 강도를 나타내지는 않으나, 특정 감성의 언급 여부를 알 수 있으므로 유사도 점수보다 적절한 결과를 도출할 수 있을 것이라고 생각하였다.

먼저 2020년 11월 10일에 가장 많이 도출된 명사 단어는 '시작'으로 179번 언급된 것을 알 수 있다. '시작'은 기쁨 차원에 속하는 감정단어이며, 유사도 점수 0.422점을 가진다. 따라서 '시작'과 관련된 유사도 점수를 계산해보면 $179 \times 0.422 = 75.538$ 이다. 동사 단어에서는 슬픔 차원에 속하는 '끝나' 단어가, 형용사 단어에서는 분노와 혐오에 속하는 '아니'라는 단어가 다른 단어들에 비해 월등히 높은 빈도를 보이는 것으로 나타났다. 12월에는 기대 차원과 즐거움, 믿음 차원에 속하는 '치료제'와 혐오, 공포, 슬픔 차원에 속하는 '환

자'가 각각 821회, 581회로 가장 높은 명사 단어 빈도를 보였으며 동사 단어에서는 '끝나'가, 명사 단어에서는 '아니'가 가장 많이 표현되었다. 더하여 앞의 11월보다 확진자 수가 폭발적으로 증가함에 따라 단어들의 빈도도 함께 전체적으로 증가했음을 알 수 있다.

2021년 1월 10일의 상위 빈출 단어는 공포 차원 단어인 '감염'이 836회, '끝나'가 336회, '아니'가 607회로 다른 단어들보다 압도적으로 많이 표현되었으며, 2월에는 '감염', '끝나', '아니' 단어가 가장 빈번하게 표현되었다. 3월에는 '치료제', '끝나', '아니' 단어가, 4월에는 '환자', '끝나' '아니'가, 5월에는 '치료제', '끝나', '아니'가 가장 상위 빈출단어로 나타났다.

백신 접종과 관련하여 다양한 이슈가 많았던 6월에는 기대 차원의 '접종'이 501회로 가장 많이 등장하였으며, '끝나'와 '아니'가 각 상위의 동사-형용사 단어로 나타났다. 확진자 수가 다시 급등한 7월에는 분노 차원의 '조심'이라는 명사 단어가 압도적으로 높게 나타났으며 동사 단어에서는 공포 차원의 '걸리'와, 공포와 기대 차원의 '기다리'라는 단어가, 형용사 단어에서는 '아니'과 즐거움 차원의 '고맙'이라는 단어가 가장 높게 나타났다. 8월에는 슬픔 차원의 '일탈'과 '끝나', '아니'가 가장 많이 등장하였다. 9월에는 '환자'와 '끝나', '아니'가, 마지막으로 10월에는 '일탈' '끝나' '아니'가 각각 상위 빈출 단어로 나타났다. 이 외에도 분석을 통해 도출된 상위 빈출 명사 단어에는 백신, 마스크와 같은 단어도 많이 등장하였으나, 구축된 감성사전에는 단어 및 유사도 점수 기준이 없으므로 분석이 불가능하여 제외하였다.

분석된 결과를 종합적으로 보면, 월별 상위 빈출 단어의 차이는 크게 존재하지 않았다. 즉, 트위터에서 빈번하게 표현하는 코로나 관련 단어들은 매 달 거의 비슷하게 구성되어 있음을 알 수 있다. 명사 단어는 대개 '치료제', '환자', '감염' '검사'와 같은 단어들이 도출되었으며 동사 형태로는 '끝나', '걸리', '지나', '미치'와 같은 단어들이 표현되었다. 형용사에서는 모든 월에서 '아니'라는 단어가 압도적인 빈출 횟수를 보였으며 이 외에 '안되', '심화', '무섭', '아프'와 같은 표현이 대거 등장하였다. 또한 세 품사의 상위 빈도 단어들 거의 부정적 차원의 감정에 속하는 단어들로 구성되어 있음을 파악할 수 있었다.

V. 결 론

본 연구에서는 장기화되는 국내 코로나 상황에 따라 소셜 미디어에서 나타나는 공중의 심리상태는 어떠한지 살펴보고자 하였다. 이를 위해 대표적인 소셜 미디어 플랫폼인 트위터를 활용하였으며, 2020년 11월부터 2021년 10월까지 약 1년 동안 수집된 자료를 대상으로 연구를 진행하였다.

분석결과 코로나 상황의 변화가 트위터에서 감정을 표출하

는 데 있어 공중에게 영향을 미치고 있음을 확인할 수 있었다. 확진자가 폭발적으로 증가했던 3차 대유행 시기와 4차 대유행 시기에서는 각각 '분노'와 '공포'가 상위로 느끼는 감정으로 나타났다. 반면에 백신 예방접종 계획이 본격적으로 수립된 시점인 2021년 1~2월과 백신 예방접종이 막 시작된 3월은 '기대'와 '신뢰'의 감정을 가장 많이 느끼는 것으로 나타났다. 이러한 결과로 볼 때, 공중이 트위터에서 커뮤니케이션을 통해 감정을 표현하는 데 있어 시기별 이슈를 반영하고 있음을 알 수 있었다.

그러나 백신 예방접종 시작에 따른 높은 '기대'와 '신뢰'에도 불구하고 공중은 '공포'와 '분노'와 같은 상반된 감정을 함께 느끼는 것으로 나타났다. 이는 불확실한 팬데믹 상황에서 공중에게 다양한 이유로 부정적 감정이 표출되고 있음을 알 수 있는 결과이다. 더하여 지난 1년간 코로나 관련 트위터에서 나타난 감성 유사도 점수를 살펴보면 '공포'와 '분노'가 가장 높은 감성 유사도 점수를 보이는 것으로 나타났다. 이러한 결과는 부정적 감정이 사회 전반에 깔려있음을 뒷받침하는 근거로 보인다.

확진자가 증가하는 3차 대유행과 4차 대유행 시작 시점에는 여덟 차원에 대한 감정 표현 전체가 다른 기간보다는 확연히 증가함을 확인하였다. 흥미로운 점은 3차 대유행 때보다 4차 대유행 시기에서 수집된 총 트위터의 양은 증가하였으나 감정의 표현은 오히려 감소하였다는 것이다. 이러한 결과는 장기화된 코로나 상황과 증가하는 확진자 수에 익숙해진 사람들이 감정 표현에 다소 무뎠어 나타난 결과로 판단된다. 그러나 감정 표현이 줄어든 가운데 '분노'와 '공포'에 관한 감정 표현은 여전히 상위 빈출 감정으로 나타난다. 따라서 이러한 부정적 감정이 다시 증가할 가능성에 대한 경계가 필요할 것으로 사료된다.

본 연구가 갖는 학술적 의의는 다음과 같다. 첫째, 본 연구에서는 극성(긍정/부정)을 넘어 여덟 차원(분노/기대/혐오/공포/기쁨/슬픔/놀람/민음)의 감성분석을 시도하여 보다 구체적인 텍스트의 특성을 파악하고자 시도하였다는 점이다.

둘째, 본 연구에서는 R 프로그램을 활용하여 약 1천만여 개에 달하는 트위터를 수집하고, 이를 통해 코로나 상황에 대한 공중의 심리상태와 주요 관심사를 실증적으로 살펴보았다. 점에서 학술적 의의를 갖는다.

본 연구에서는 다음과 같은 한계점을 가진다. 첫째, 본 연구는 트위터 데이터를 대상으로 분석한 것임을 감안할 때, 도출된 결과가 전 연령층을 포괄했다고 보기에는 다소 무리가 있을 것으로 생각된다. 트위터에 따르면 서비스를 사용하는 주 연령층은 젊은 층이 대부분인데, 실제로 사용자의 57.1%가 25세에서 49세 사이인 것으로 나타났다.

둘째, 감염병 상황에서의 감성분석을 위해 보다 면밀한 감성사전의 구축이 필요할 것으로 사료된다. 감성분석을 실시하기에 앞서 사전의 정제와 보완을 수행하였지만 방대한 양의 사전은 소수의 연구자가 완전한 형태로 구성하기에는 부족함이 있었을 것으로 추정된다.

감사의 글

이 논문은 2021년도 한림대학교 교비연구비(HRF-202106-003)에 의하여 연구되었음.

참고문헌

- [1] M. S. Ann, "A Public Perception Study on the new word "Corona Blue": Focusing on Social Media Big Data Analysis," *International Journal of Advanced Culture Technology*, Vol. 8, No. 3, pp. 133-139, September 2020. <https://doi.org/10.17703/IJACT.2020.8.3.133>
- [2] Y. Lim, "COVID-19 Blues: A Big Data Analysis," *Korean Journal of Counseling and Psychotherapy*, Vol. 33, No. 2, pp. 829-852, 2021.
- [3] E. Lee, COVID-19 Generation. How's your Mental Health!, GRI, No. 414, 2020.
- [4] J. Liu, Y. Bao, X. Huang, J. Shi, and L. Lu, "Mental Health Considerations for Children Quarantined Because of COVID-19," *The Lancet Child & Adolescent Health*, Vol. 4, No. 5, pp. 347-349, May 2020. [https://doi.org/10.1016/S2352-4642\(20\)30096-1](https://doi.org/10.1016/S2352-4642(20)30096-1)
- [5] H. Son, K. Kim, J. Huh, and C. Park, Busan City Policy Response to Overcome COVID-19, BDI Policy Focus, pp. 1-12, 2020.
- [6] M. Taylor, G. Wells, G. Howell, and B. Raphael, "The Role of Social Media as Psychological First Aid as a Support To Community Resilience Building," *The Australian Journal of Emergency Management*, Vol. 27, No. 1, pp. 20-26, 2012.
- [7] N. Athanasis, M. Themistocleous, K. Kalabokidis, A. Papakonstantinou, N. Soulakellis, and P. Palaiologou, "The Emergence of Social Media for Natural Disasters Management: A Big Data Perspective," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 42, No. 3, pp. 75-82, March 2018. <https://doi.org/10.5194/ISPRS-ARCHIVES-XLII-3-W4-75-2018>
- [8] S. Ham, *Social Media and Crisis Management PR*. Seoul: Communication Books, 2013.
- [9] O. Phelan, K. McCarthy, and B. Smyth, "Using Twitter To Recommend Real-time Topical News," in *Proceedings of the 3rd ACM Conference on Recommender Systems*, New York, USA, pp. 385-388, 2009. <https://doi.org/10.1145/1639714.1639794>
- [10] S. Lee, S. Ryu, and S. Ahn, "Mass Media and Social Media Agenda Analysis Using Text Mining: Focused on '5-day Rotation Mask Distribution System'," *The Journal of the Korea Contents Association*, Vol. 20, No. 6, pp. 460-469, June 2020. <https://doi.org/10.5392/JKCA.2020.20.06.460>
- [11] S. Cresci, A. Cimino, F. Dell'Orletta, and M. Tesconi, "Crisis Mapping During Natural Disasters via Text Analysis of Social Media Messages," in *Proceedings of International Conference on Web Information Systems Engineering*, Miami, USA, pp. 250-258, 2015. https://doi.org/10.1007/978-3-319-26187-4_21
- [12] J. Oh and K. Jung, "Risk Communication Network Analysis Based on Disaster Response Framework: Focusing on the 2017 Gangneung and Samcheok forest fires," in *Proceedings of 2017 The Korean Association for Policy Studies Conference*, Seoul, Korea, Vol. 2017, No. 1, pp. 128-141, 2017.
- [13] S. Kim, D. Kim, and M. H. Ryu, "Crisis Communication on Social Media during COVID-19 Pandemic: An Analysis of Facebook and YouTube," *Journal of the Korea Industrial Information Systems Research*, Vol. 26, No. 6, pp. 47-60, December 2021. <http://dx.doi.org/10.9723/jksis.2021.26.6.047>
- [14] A. Hotho, A. Nürnberger, and G. Paaß, "A Brief Survey of Text Mining," *In Ldv Forum*, Vol. 20, No. 1, pp. 19-62, 2005.
- [15] K. Cho and N. Ha, "Research Trends on Emotional Labor in Korea Using Text Mining," *Journal of the Korea Industrial Information Systems Research*, Vol. 26, No. 6, pp. 119-133, December 2021. <http://dx.doi.org/10.9723/jksis.2021.26.6.119>
- [16] D. Choi, B. Song, D. Park, and S. Lee, "Keyword Trends Analysis Related to the Aviation Industry During the Covid-19 Period Using Text Mining," *Journal of the Korea Industrial Information Systems Research*, Vol. 27, No. 2, pp. 115-128, 2022.
- [17] W. K. Chen, D. Riantama, and L. S. Chen, "Using a Text Mining Approach To Hear Voices of Customers from Social Media toward the Fast-food Restaurant Industry," *Sustainability*, Vol. 13, No. 1, 268, December 2021. <https://doi.org/10.3390/su13010268>
- [18] W. Deng, J. H. Hsu, K. Löfgren, and W. Cho, "Who Is Leading China's Family Planning Policy Discourse in Weibo? A Social Media Text Mining Analysis," *Policy & Internet*, Vol. 13, No. 4, pp. 485-501, December 2021. <https://doi.org/10.1002/poi3.264>
- [19] N. Öztürk and S. Ayvaz, "Sentiment Analysis on Twitter: A Text Mining Approach to the Syrian Refugee Crisis," *Telematics and Informatics*, Vol. 35, No. 1, pp. 136-147, April 2018. <https://doi.org/10.1016/j.tele.2017.10.006>

- [20] E. Dubois and D. Gaffney, "The Multiple Facets of Influence: Identifying Political Influentials and Opinion Leaders on Twitter," *American Behavioral Scientist*, Vol. 58, No. 10, pp. 1260-1277, April 2014.
<https://doi.org/10.1177/0002764214527088>
- [21] W. W. Xu, Y. Sang, S. Blasiola, H., and W. Park, "Predicting Opinion Leaders in Twitter Activism Networks: The Case of the Wisconsin Recall Election," *American Behavioral Scientist*, Vol. 58, No. 10, pp. 1278-1293. March 2014.
<https://doi.org/10.1177/0002764214527091>



용혜련 (Hye-Ryeon Yong)

2017년 : 한림대학교 인터랙션디자인 대학원 (인터랙션디자인학 석사)
2022년 : 한림대학교 인터랙션디자인 대학원 (인터랙션디자인학 박사)

2022년~현재 : 효성ITX

※ 관심분야 : 인터랙션 디자인(Interaction Design), 빅데이터 분석(Big Data Analytics)



황현석 (Hyun-Seok Hwang)

2000년 : POSTECH 대학원 (공학석사-경영정보시스템)
2004년 : POSTECH 대학원 (공학박사-경영정보시스템)

2004년~현재 : 한림대학교 경영학과 교수

※ 관심분야 : 경영정보시스템(Management Information System), 데이터 사이언스(Data Science), 빅데이터 분석(Big Data Analytics), 데이터 분석